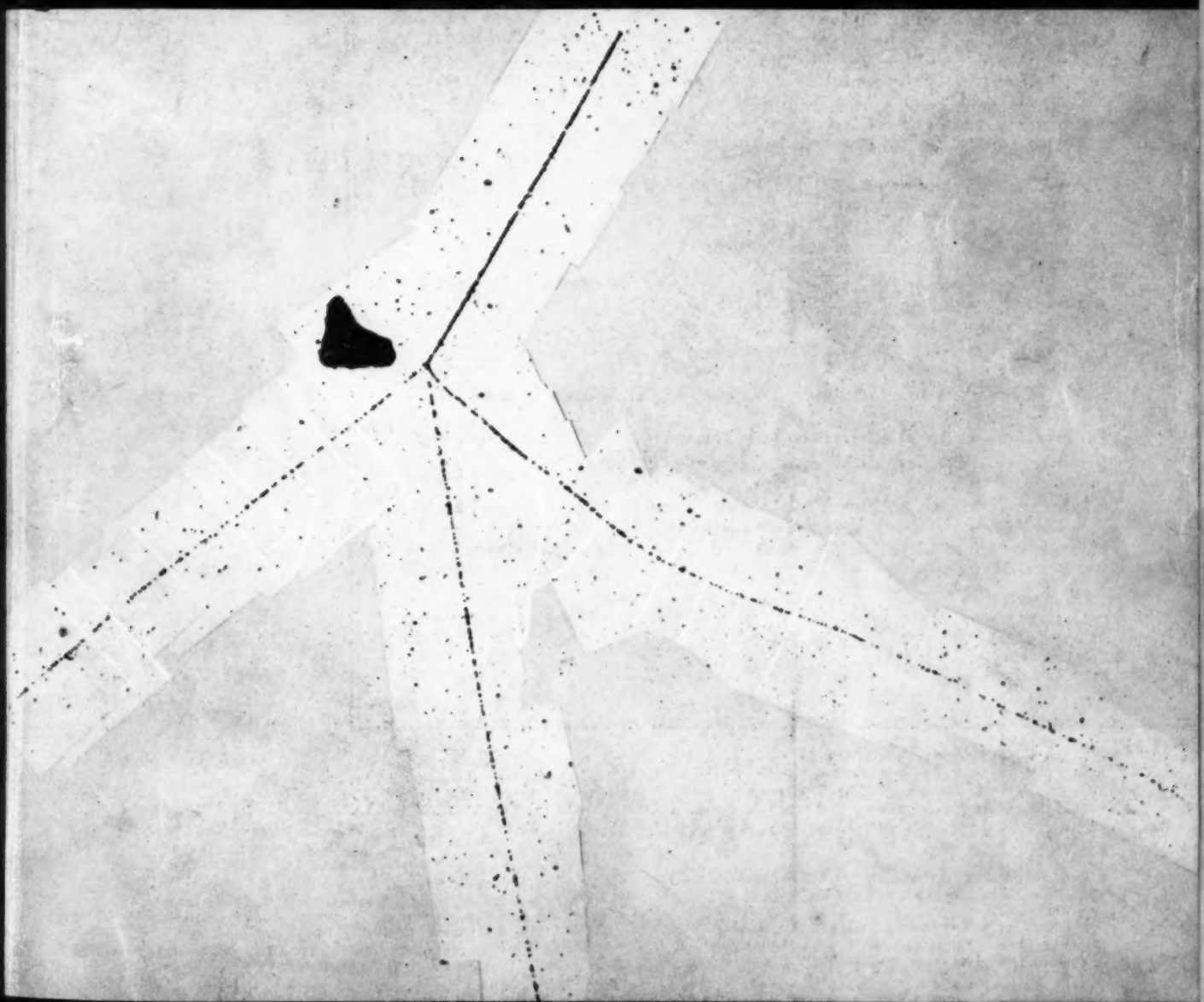
JOURNAL OF APPLIED PHYSICS



JOURNAL
OF
APPLIED PHYSICS

Vol. 19, No. 4 APRIL, 1948

In This Issue

		Page
Physics in 1947	Philip Morrison	311

Contributed Original Research

Self-Excited Mechanical Oscillations	N. Minorsky	332
Erratum: A Magnetron Resonator System	E. C. Okress	338
Electronic Simultaneous Equation Solver	Edwin A. Goldberg and George W. Brown	339
Surface Waves from an Underwater Explosion	J. G. Kirkwood and R. J. Seeger	346
Calculation of the Temperature Development in a Contact Heated in the Contact Surface, and Application to the Problem of the Temperature Rise in a Sliding Contact	Ragnar Holm	361
Measurement of Carbon Black Particles by the Electron Microscope and Low Temperature Nitrogen Adsorption Isotherms	R. B. Anderson and P. H. Emmett	367
Visco-Elastic Behavior of a Highly Plasticized Nitrocellulose in Compression under Constant Load	George J. Doyle and Richard M. Badger	373
Oxide Replica Technique for the Electron Microscope Examination of Stainless Steel and High Nickel Alloys	E. M. Mahla and N. A. Nielson	378
Head-On Collision of a Shock Wave and a Rarefaction Wave in One Dimension	H. E. Moses	383
Preferred Orientation Determination Using a Geiger Counter X-Ray Diffraction Goniometer	B. F. Decker, E. T. Asp, and D. Harker	388
Effect of Radiation on the Rate of Burning of Solid Fuel Rocket Propellants	S. S. Penner	392
Observations on Arc Discharges at Low Pressures	M. J. Kofoid	399
Stresses in a Plate Containing an Overlapped Circular Hole	Chih-Bing Ling	405
Minimum Detectable Absorption in Microwave Spectroscopy and an Analysis of the Stark Modulation Method	W. D. Hershberger	411
Measurements of Dielectric Constant and Dipole Moment of Gases by the Beat-Frequency Method	J. G. Jelatis	419

In Every Issue

Calendar of Meetings	331
Here and There	425
New Booklets	426
Index to Advertisers	xliv

BOARD OF EDITORS

Elmer Hutchisson,
Editor

Associate Editors

R. Bowling Barnes
Otto Beeck
W. F. Busse
W. James Lyons
Philip Morrison
Paul E. Sabine
Frederick Seitz
J. R. Townsend
John Von Neumann

Publication of the American Institute of Physics

ADMINISTRATIVE STAFF

Henry A. Barton,
Director

Cleveland Norcross,
Assistant Director

Ruth F. Bryans,
Publication Manager

Theodore Vorburger,
Business Manager

John T. Tate,
Adviser on Publications

GOVERNING BOARD

George R. Harrison,
Chairman

Wallace Waterfall,
Secretary

George B. Pegram,
Treasurer

Luis W. Alvarez
Robert B. Brode
E. U. Condon
A. J. Dempster
W. F. Fair, Jr.
Harvey Fletcher
R. C. Gibbs
W. V. Houston
Paul E. Klopsteg
F. W. Loomis
Philip M. Morse
A. H. Pfund
John T. Tate
M. W. White
A. G. Worthing

The Journal of Applied Physics, published monthly at Prince and Lemon Sts., Lancaster, Pa., is devoted to physics in its role as the science basic to other natural sciences and to the arts and industries. Previous to 1937 this publication was known as Physics. The Journal of Applied Physics publishes editorials, news of physicists, and reviews, as well as technical papers of applied physics.

Manuscripts should be submitted to Elmer Hutchisson, Editor, Case Institute of Technology, Cleveland 6, Ohio.

Subscriptions and orders for back numbers should be addressed to Prince and Lemon Streets, Lancaster, Pa. or to the American Institute of Physics, 57 East 55 Street, New York 22, N. Y.

Subscription Price	U.S. and Canada	Elsewhere
To members and associate members of the American Institute of Physics	$6.00	$7.00
To all others	8.00	9.00

Back Number Prices

Complete set: Vol. 1, July, 1931–Vol. 18, 1947—$132.00
Yearly back number rate: $10.00
Single copies: $1.00 each.

Changes of address should be sent to the Business Manager, American Institute of Physics; **corrected proofs** should be sent to the Publication Manager, American Institute of Physics.

Advertising rates supplied on request. Orders, advertising copy, and cuts should be sent to the American Institute of Physics.

Entered as second class matter January 22, 1937, at the Post Office at Lancaster, Pennsylvania, under the Act of March 3, 1879. Accepted for mailing at the special rate of postage provided for in the Act of February 28, 1925, authorized May 2, 1932.

Copyright 1948, by the American Institute of Physics.

Journal
of
Applied Physics

Volume 19, Number 4 April, 1948

Physics in 1947

Philip Morrison
Cornell University, Ithaca, New York
(Received February 24, 1948)

LABORATORIES in 1947 had settled down. There were still problems, serious ones both for physicists and for Everyman, but no one was looking for rapid solutions. The shuffling of men to new jobs, the hard decisions, and the hasty changes, were no longer commonplace. The plans were beginning to take shape. The great machines were turning from budget items and drawings into rolled plates and banks of condensers, still unwired. But even without them, progress was in the air. The year brought great results, and seemed to bring physics to the threshold of a real change. It was a year like, perhaps, 1924 or 1931, though not yet like the years of harvest which followed, when the wave mechanics and the nuclear particles were made part of the science. It was a year of excitement and discovery. And 1948 lies ahead!

Not the least satisfying part of the year's work was the re-establishment, in part, of the international community of science. Many Americans went abroad, and there were no small numbers of visitors here. Best of all, many physicists abroad, especially the British and the workers at Rome, produced results which demanded world-wide attention. Once again you hurried to the library to see the latest copy of Nature. It was almost like old times, though there was still a long way to go.

This account will try to survey the year in physics. It will be clear that physics is too diverse to be caught in a net of so coarse a weave; perhaps it would be fairer to say that the survey will be of advances in some of the more unsettled branches of physics, in the work in nuclei and the fundamental particles, in cryogeny and radioactivity, and cosmic rays. There is much that is good and exciting that must be left out; but there is plenty to record.

THE UNFUNDAMENTAL PARTICLES

The classical elements were earth, air, fire, and water. The rise of modern science brought a more sophisticated but less satisfying tale, and the ninety-six elements we now count exhaust our list of atomic species. But since Prout there has been a disposition to unify matter by regarding the elements as not fundamental, except for the chemists, and to find the blocks from which they were built. The neutron and the proton seem adequate to form nuclei; the electron and its twin, the positron, enable us to make all matter; the neutrino satisfies the conservation laws for energy and momentum in that curious phenomenon, beta-decay; and the photon, the quantum of the electromagnetic field, serves to describe the phenomena of Maxwell's equation in the era of quantum-mechanics. The atoms and their aggregates, molecules and crystals, are held together by wholly electrical forces associated with the photons, by now familiar. This is not too long a list, though not as simple as the ancients wished.

Since 1936, when Professor Yukawa of Kyoto first suggested that a heavy quantum would be adequate to form a field capable of holding together the nuclear particles, the meson has been on the list. When, in the years after Yukawa's suggestion, the penetrating component of the cosmic ray was shown indeed to be a particle of mass about 200 times the electron mass, occurring both with positive and negative charge, the meson was identified. But a complete description such as we have for electrical forces has not been given to us for the "mesonic" forces. These forces, in which the neutral or electrically-charged mesons play the role of photons, and the nuclear particles the role of sources of these heavy quanta, are still not known in detail, much less understood. We are in the stage before the inverse-square law; about all we know is a somewhat more quantitative analog to "unlikes attract; likes repel." In this state theory has tried to make generalizations. In more than one of these, the list of fundamental particles has been increased by hypothesizing a neutral meson, like the charged ones we see, and even a whole trinity of positive, negative, and neutral, but of a different mass. In 1947 at least three and perhaps as many as five new "mesons" have been added to the tally not as speculation, but by unimpeachable observations of nature. The word "fundamental," always implying a tacit qualification, is now pretty dilute; "unfundamental" would be about as happy. It is far too soon to say, but it is likely that the next years will see the number of "fundamental" particles multiply, until, one may hope, a new and unifying concept will again make it possible to use the word proudly.

It is the biggest story of the year, perhaps, the story of how the new mesons were found. It was not with the great machines, but with that inexhaustible source of surprises, the cosmic ray.

At sea level most of the ionizing particles of cosmic rays are mesons, about six-tenths of them

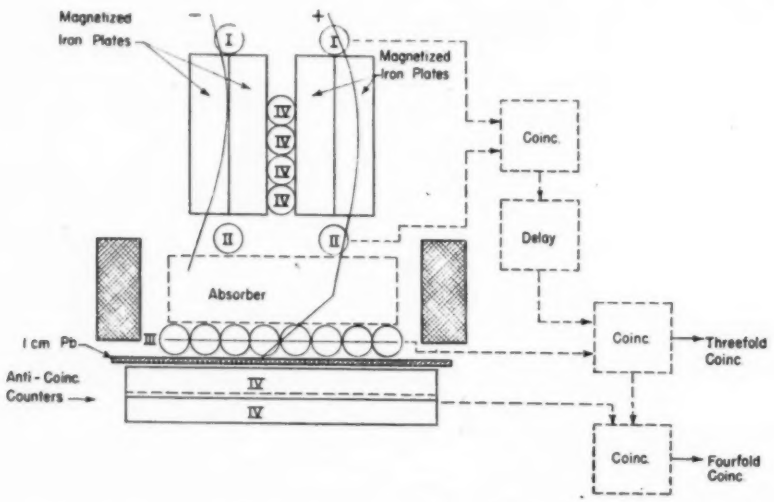


FIG. 1. A diagram of the Rome apparatus, supplied by Dr. Piccioni. A meson enters from the top, crossing one of the counters I. The iron plates are magnetized in such a direction that positive charges would be concentrated somewhat at counters II, while negative ones would be dispersed. This is indicated in the diagram. Such particles will traverse counters II. If they then decay in the large block of absorber, after a delay of some microseconds, which can be varied by the experimenter, and emit an electron to count in counters III, the proper event has been recorded. If any of counters IV record, it is then clear either that the ray did not enter the apparatus alone, or that it produced secondaries on its way through. Both of these must be rejected; mesons do not produce secondaries in this path length, except by decay. The lead at the sides shields the counters against sidewise—moving rays, and the thin lead sheet below stops the decay electron before it can actuate the counter set below. A count is then taken of the rays which count in I, in II, and in III a few microseconds later, but not in IV. The number of such counts is compared for reversal of magnetic field, and for different absorber materials. The results are described in the text. This is a typical cosmic-ray counter experiment, susceptible of many variations.

positive, the rest negative in charge. These penetrate matter well, because they are heavy compared to electrons, and lose momentum less easily in the glancing collisions with the electrons of the atoms they pass. But they interacted strongly with nuclei, it was held, not by electrical forces alone, but by the specific mesonic force, the same which held the nucleus together. To be sure, a few experiments had been done to see how large was the interaction of these mesons with nuclei. It did not appear large, but the effect was superimposed on the electrical scattering, and one might doubt the experiment. The penetration of these rays to great depths also seemed difficult to understand. Why did they not get stopped more quickly by nuclei? There has always been a difficulty here for the experts.

Consider a meson coming into a slab of iron. It will lose energy rapidly, frittering it away mainly by ionization, and if it is not too energetic to begin with, it will come to rest in the iron slab. What then? We do not find mesons lying on the laboratory floor. There was a double answer to this. For negative mesons, it was pointed out years ago that they would come to rest, be caught in a stable orbit deep inside an atom, held there by the attraction of the positive nucleus, until their strong specific interaction with nuclear matter would cause their absorption and a nuclear transformation of some kind, perhaps a so-called "star" disintegration, in which several nuclear fragments fly out of the nucleus, their cloud-chamber tracks making the rays of a rough many-pointed star. The positive mesons which stopped in matter would have a different fate. They are repelled by nuclei and, once at rest, could not be appreciably absorbed by the nucleus at a large distance—a few angstroms—since the nuclear forces are so very short-ranged. But neither would they last forever. For it was known that mesons fly apart in flight; more conventionally, that they undergo spontaneous disintegration in free space, with a probability proportional only to the time. In short, these mesons are radioactive; they beta-decay, releasing electron or positron, depending on the original sign of charge. For this there were cloud-chamber tracks in proof, showing a meson coming to rest and an electron starting from the point where the spent meson had died. In direct cosmic-ray work, too,

FIG. 2. Two adjacent frames of Valley's photographs. The top frame shows a positive meson on its way to decay in the carbon block below the cloud chamber. The second frame shows a negative meson. These films were exposed after the meson had decayed, when the cloud chamber was expanded and the track of the decaying meson caught. The delay time was simultaneously measured. The highly curved negative track at the top of the upper picture is the result of a particle which entered the cloud chamber after the expansion. Its droplets have not diffused, and the track is very sharp.

there had been a set of measurements showing that the meson "beam" was attenuated more by a kilometer of air than by its equivalent weight of water or lead. The time taken to travel the distance in space "absorbed" mesons. Even the lifetime was known, about two microseconds.

The year began with a surprise. It was the report from Rome of work done there, begun during the German occupation, and carried out under conditions which would have caused any but the most enthusiastic to falter. Two young physicists of the Rome group, Conversi and Piccioni, had studied the decay of mesons stopped in their apparatus. The arrangement is shown and the details of the experiment described in Fig. 1 and

its caption.[1] They found that indeed the positive mesons decayed in iron, while the negative ones did not and were presumably captured before they had a chance to decay. But in carbon, nearly all the mesons decayed; few could be captured by nuclei. The mesons do *not* interact strongly with nuclei.

This work has been confirmed in detail during the year. Figure 2 shows photographs taken by the M.I.T. group, Valley and Rossi, who have made a beautiful and direct check[2] of the pioneer work of the Rome experimenters. In this experiment, mesons of both signs are allowed to enter the apparatus (see Fig. 3). But when after a decent wait of a few microseconds a delayed decay of a stopped meson has been recorded, a cloud-chamber expansion is initiated. In this way the track of the meson which had decayed is made visible, and the magnetic field across the chamber allows unambiguous determination of the sign of the particle. Valley proved that in steel and brass, positive mesons decay, while the negatives are captured. In carbon, beryllium, and water, capture is so improbable a process that even the negatives live out their natural lives and decay. The influence of the capture probability on the apparent lifetime is not by any means negligible, and shows up as an effective reduction of the lifetime of negative mesons. In a material like aluminum or magnesium, the chance that a negative meson will decay on its own is quite comparable to the chance that it will be swallowed up by a nucleus.

The theorists—and the builders of machines—were worried. Mesons had been found as Yukawa had guessed. But they did not interact with nuclei. Did that mean that they had nothing to do with nuclear forces? Mesons were somehow produced from the nuclei of the upper air; they were too short-lived to be primary cosmic-ray particles. If they could be produced in a single step, they must be directly absorbable. There was one untested link in the argument. The theorists had to strengthen it.[3] It was the tacit assumption that the atomic process which slowed mesons down and then brought the negative ones into a deep-lying "orbit," far within the atomic electron shell, where they frequently collided directly with the nucleus, was a fast process. Perhaps the delay was only here, and the mesons were capturable, but never got close enough before they died of beta-decay. This argument was put forward by some of the cagiest physicists, but was disposed of by detailed calculation. It turns out that the meson is slowed down from high speed to a velocity where its capture into an atomic orbit is possible in about 10^{-9} second at most, and that its further loss of energy to fall into its K-shell orbit—only about 10^{-12} cm in radius—is very rapid. In an additional time of less than 10^{-12} second after it has been captured into any atomic orbit, the meson has lost momentum enough to the electron cloud which whirls by it so that it falls to the innermost orbit. Then the nucleus can capture it, if that process is allowed. The fact that nuclei capture the mesons so slowly must then be ascribed to a relatively weak interaction, perhaps 10^8 times weaker than that predicted quite generally by theories which used these mesons to explain the nuclear forces. Here was a real crisis. Moreover, if mesons were not captured easily, neither could they be produced easily. The great machines might make no mesons.

And yet they were made, nevertheless. High in the air the primary cosmic ray made mesons freely. The contradiction was resolved by some sensible speculation, and really explained by simultaneous beautiful experiment. The idea which lay behind the explanation was this: however reluctantly, we had to postulate another kind of meson. The new meson could be made easily from nuclei; it would be captured rapidly. But the meson we see at sea level is *not* the meson which interacts to produce the forces between neutron and proton. It is the decay product of the new meson, spontaneously produced by a special process from the postulated new particle. At the price of unsupported speculation about new particles, sense could be made from the paradox between easy manufacture of mesons in the upper air, difficult capture of mesons here in the apparatus of Piccioni and Valley, and the others. This

[1] Conversi, Pancini, and Piccioni, Phys. Rev. **71**, 209 (1947).
[2] G. E. Valley, Phys. Rev. **72**, 772 (1947); G. E. Valley and B. Rossi, Phys. Rev. **73**, 177 (1948).
[3] E. Fermi and E. Teller, Phys. Rev. **72**, 399 (1947).

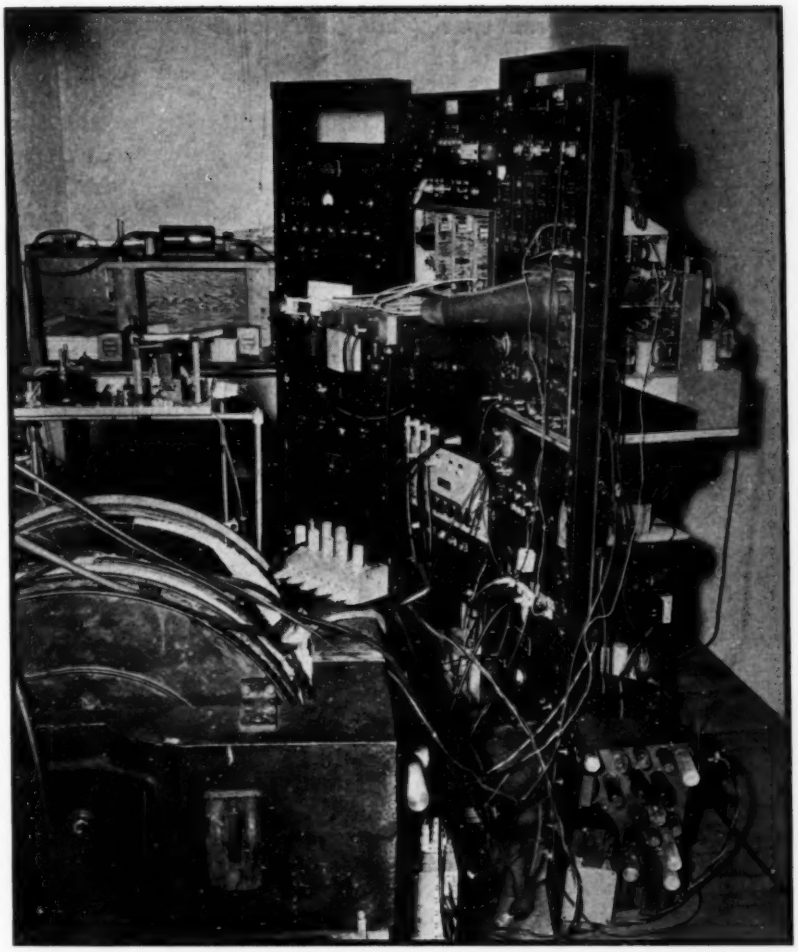


FIG. 3. The equipment of Valley and Rossi at M.I.T. Professor Valley's photograph shows the large magnet used to curve the tracks of mesons in the cloud chamber in the left foreground. Behind the magnet is the black box holding the camera. A small horizontal shelf holds a microsecond time meter which faces downward. The camera photographs the track and the time elapsing before the meson which made the track decayed in the absorber below the cloud chamber. The relay racks at the right hold the electronic equipment for cloud-chamber, counter-coincidence, and time-delay measurement.

tentative suggestion formed the basis for what has been called the two-meson hypothesis.[4]

Almost before these notions had been formed, their most daring assumptions turned into rather evident facts. The cosmic-ray group at the University of Bristol, led by Powell, had found a brand-new meson, or maybe two of them.[5] These new particles gave plenty of leeway for rationalizing the difficulties found so far. The Bristol experiments are simple, powerful, and elegant. They use no complicated series of coincidence circuits, no microsecond timing circuits, no magnets. They employ a photographic plate, a microscope, and plenty of insight, patience, and skill, always the best of apparatus.

[4] R. E. Marshak and H. A. Bethe, Phys. Rev. **72**, 506 (1947). A similar idea is reported in Japanese journals by Sakata and others in 1943.

[5] Lattes, Powell, and Occhialini, Nature **160**, 453, 486 (1947).

Microscopic examination of an unexposed but developed photographic plate always shows a more or less random distribution of blackened grains. A grain of photographic emulsion can be made developable by the production within it of a number of excited or ionized atoms. Thus if a fast-charged particle traverses the plate, its track will be marked exactly as in a cloud chamber by a visible trail of developed grains when the plate is treated to develop the latent image. These tracks are easily identified, as Figs. 4 and 5 show, by their characteristic appearance. The number of ions required to render a grain developable will depend upon the properties of the emulsion used. A single gamma-ray or fast electron makes so few ions in crossing a single grain that as yet no plate will show such a track. The blackening of film by gamma-rays requires an intensity so great that several secondary electrons cross a single grain of

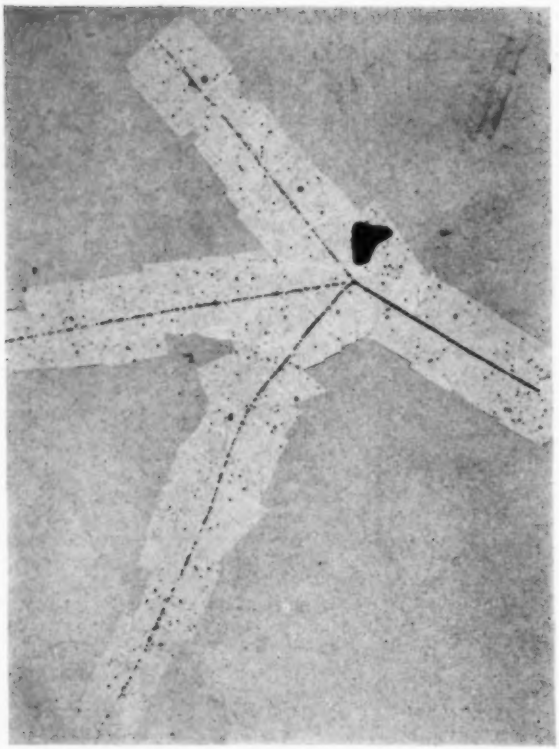


FIG. 4. One of the microphotographs of the plates of the Bristol group. The picture reproduced is a mosaic of some thirty-odd different microphotographs, the microscope being refocused and moved for each exposure because of the small field and depth of focus compared to the track length. This whole event occupies less than a twentieth of a millimeter in the original emulsion. The long curved track, which shows an increasing density of grains as it becomes slower and piles up considerable curvature as a result of multiple scattering encounters, is a slow meson, of Powell's new type, ending in the emulsion. This meson is absorbed by a nucleus, which disintegrates, emitting three visible heavy particles, in a cosmic-ray star process.

the sensitive material. But the heavily ionizing tracks of protons will show up clearly in almost any properly made film. This effect has been used for years by many men, especially by the workers of the Bristol group.

In the last years the Bristol group, cooperating with the great English photographic manufacturers, Ilford, have produced plates of great sensitivity. Here the ionization density required to produce a track has been cut sharply, not indeed enough to make electrons visible, but far enough to make visible the tracks of slow mesons, which ionize much more weakly than the protons commonly observed. These plates were carefully calibrated, so that the Bristol workers knew the grain spacing and the range characteristic of particles of all energies of interest.[6]

The photographic plate is the equivalent of a frozen cloud chamber. It has the advantage of that unique instrument of showing the actual path of particles and the course of events, not of simply recording clicks in one of many complex channels for electronic pulse signals. Interpretation is direct. But the plate is not gaseous; it is solid. It compresses the cloud chamber a 1000-fold, and increases by that factor the chance of an interesting particle stopping in the chamber. Even more, it is sensitive for months, not just for the 0.1 second of a cloud-chamber expansion, which cannot be repeated except on a rather slow cycle. One pays for this, of course. The photographic plate must be examined over its whole surface and throughout the 100-μ depth of its emulsion by a rather high power microscope, with its small field area and limited depth of focus. Patience is needed, and long days of peering. The Bristol group has developed a team of more than a dozen trained observers—the fame of Mrs. Powell as such an observer is by now international—who study the plates carefully and who locate all events of interest, which are turned over to the physicists themselves for the careful measurements on which the quantitative results depend. The plates which have made history were sent to the high mountains of France and of Bolivia, placed quietly on a shelf there for a month or more, and then mailed back to Bristol. The examination which followed can be computed to be the equivalent of examining ***one billion*** photographs taken by random expansion in an ordinary cloud chamber! It is no wonder that rare and valuable results were obtained.

It would take too much space to describe the ingenious analysis which was made. Grain spacing, track tortuousness as a measure of scattering, track length measures, the statistical corrections needed for any such experiment, the identification of special kinds of "hammer" and "star" tracks as particular types of nuclear process, the corrections for fading of the image, and for the background tracks of radioactive substances—all these points speak for the inge-

[6] C. Lattes *et al.*, Nature **159**, 93, 186, 694 (1947).

nuity of the Bristol team. The most important observation can be summarized briefly. Some few hundred tracks have been identified as mesons, by their large scattering and low grain density. Of these, a few chance to end in the emulsion. The grain density increases as they slow down. But at the end of some such tracks a new particle, also a meson, starts up again, to end in the emulsion itself a few hundred microns away. Comparison of grain density and range showed that the secondary meson had a mass lower than the primary which engendered it. Consideration of the momentum changes given by the directions of the tracks showed that all such events could be accounted for as the disintegration of the first meson into *two* mesons, one which made the revived track, and another of about the same mass, but leaving no track of ions, and hence presumably neutral. The care and the detail of these experiments leave little room for doubt. If we assume that the secondary meson is the non-capturable kind we know at sea level, then the mesons of Powell and his co-workers are about 375 times as heavy as the electron. Here were plenty of particles for the theorists.

Nor was this all. In high altitude cloud-chamber flights in large airplanes, meson decay was observed by C. D. Anderson—just twice.[7] On both these pictures one can see a particle of intermediate mass, certainly a meson, but with mass not very well fixed under the experimental conditions in spite of a strong magnetic field, decay to emit a particle which is certainly a positron, with 25 or 30 Mev of energy. Only one particle is seen to come out. But both pictures show the same energy for the positron, which makes it seem likely that the disintegration is into two particles, one neutral. This is the old explanation given for such photos before. But never before has the energy of the secondary positron been well measured. It is only 30 Mev, as curvature measures show. If the meson which decays is the ordinary sea level variety, with a rest mass energy equivalent of 100 Mev (just 200 times more massive than the electron), the other 70 Mev go mainly into the rest mass of the neutral particle, now rightfully to be called a meson. It cannot be a gamma-ray or a massless neutrino, as we once thought. One more meson. Perhaps it is the same neutral meson which ac-

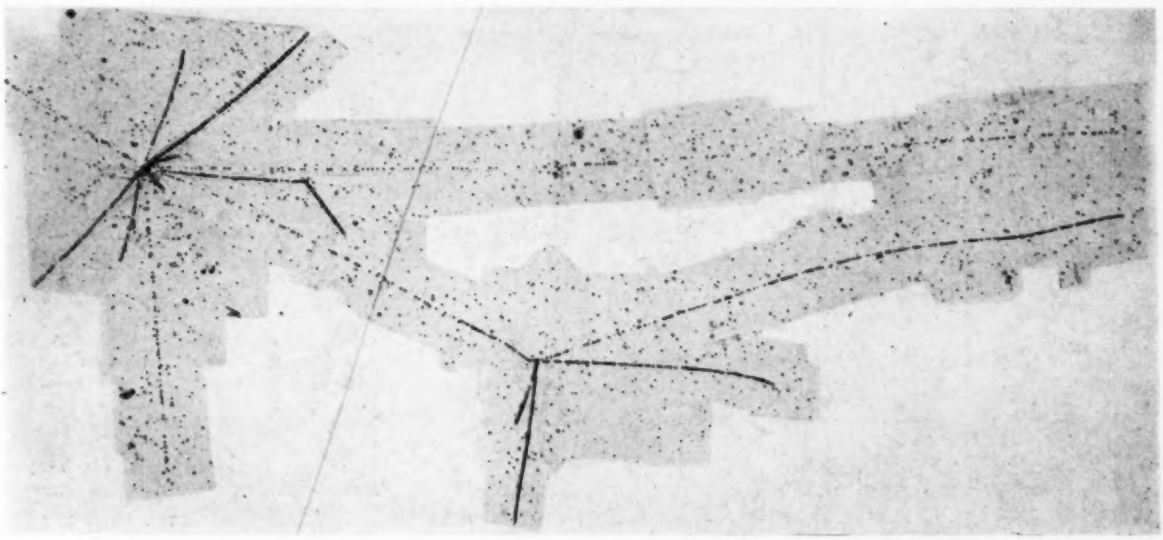


FIG. 5. A very complicated and wonderful event found in the Bristol plates. Here a nuclear star process of uncertain origin produces about a dozen charged particles. Two of these appear to be mesons. One track, certainly a meson, becomes curved and denser as it slows down, and ends by producing another star of four particles. Many more such events are observed in full detail by the workers at Bristol. Both of these photos were kindly supplied by R. A. Marshak, who had obtained them in Bristol from Powell, Lattes, and Occhialini.

[7] Anderson, Adams, Lloyd, and Rau, Phys. Rev. **72**, 724 (1947).

companies the disintegration of the Bristol heavy mesons. Such an identification is unsure, but it does lead to a rather neat theory.

Even this is not all. More cloud-chamber pictures, this time from Manchester,[8] and some mass measurements from Berkeley cosmic-ray workers,[9] and a whole series of rather indirect experiments done in the Caucasus by Russian cosmic-ray workers,[10] all point to more mesons, perhaps about a thousand times the electron mass, both charged and neutral. None of these pieces of evidence is conclusive; taken together they are at least disquieting.

The study of nuclear forces, the best use of high energy machines, the superenergy cosmic-ray phenomena, especially the giant showers—these are the problems before the workers in mesons. 1947 brought the simplest and most unsatisfactory account of mesons to a close, and entered a new era of meson sophistication. There is wide opportunity for many approaches to a partial solution of many problems. It is sure that the meson exists; and now we have a whole family of these intermediate particles, of positive, negative, and zero charge. The final unraveling of this complex and the description of nuclear forces with its aid is surely a task for a decade, even at the present rate. More mesons are not to be excluded; the unfundamental character of our fundamental particles now cries for explanation. It is certain that the reviews of the years to come will continue to illuminate this new and exciting landscape.

SPECTROSCOPY, OLD AND NEW

For a long time, now nearly thirty-five years, the atomic spectra, and especially the spectrum of the simplest atom, hydrogen, have been the touchstones of physical understanding. The Bohr theory of the atom won its success from the Balmer series; the Sommerfeld relativistic modification, from the doublet structure; and the Dirac theory, from its complete agreement with the correct Sommerfeld doublet positions, and from the intensities which it correctly gave. From the side of theory, the system was the simplest we know; from the side of experiment, the technique was enormously precise and refined. It was the place for the test of fundamental ideas.

Not all was well. In 1937 (see Fig. 6) R. C.

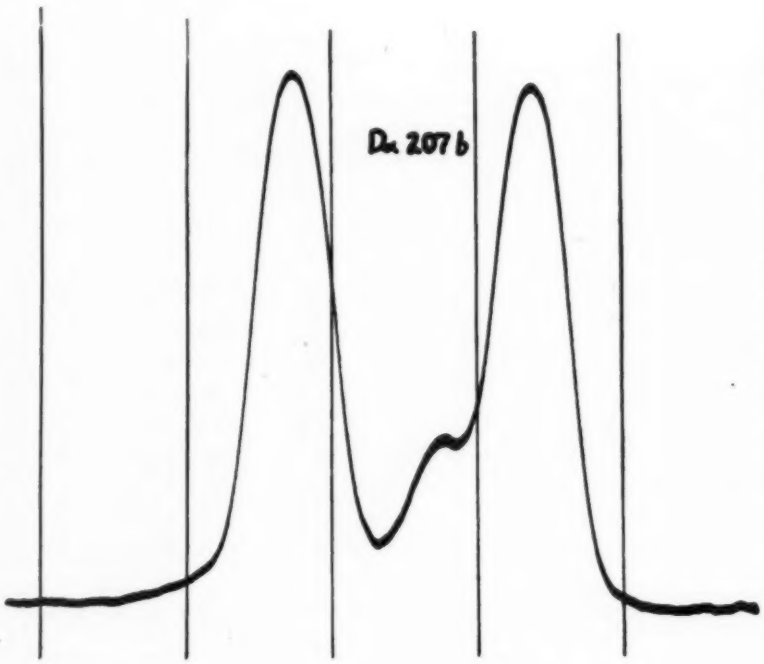


FIG. 6. This is a microphotometer tracing of a high resolution spectrograph of the strong red line of the hydrogen atom. (This is actually a photo of the D_α line in deuterium, hydrogen of weight two.) The two strong peaks are the lines of the main transitions, and the small satellite on the right-hand line is the transition from the state $3^2P_{\frac{1}{2}}$ to the state $2^2S_{\frac{1}{2}}$. According to the Dirac theory, the small peak should lie about 0.03 cm^{-1} closer to its strong neighbor than it does in this tracing. The vertical lines on the tracing are spaced about one-fifth cm^{-1}. While the line is not well resolved, it seems clear that moving it that far to the right would merge it nearly completely with the righthand doublet line. The line shift so measured was given the correct interpretation in 1938. The tracing was kindly supplied by Professor R. C. Williams; it appears in The Physical Review **54**, 561 (1938). In the same year S. Pasternack correctly evaluated the shift of the S level of hydrogen, which would explain this and other small spectroscopic effects. It took ten years to come to believe this result.

[8] Rochester *et al.*, private communication.
[9] J. G. Retallack, Phys. Rev. **72**, 742 (A) (1947).
[10] Alichanian, Alichanov, and Weissenberg, J. Phys. U.S.S.R. **11**, 97 (1947).

Williams, then at Cornell, made the best of a number of precision and high resolution spectroscopic examinations of the fine structure of the prominent deep red line of the hydrogen atom—the well-known H_α line—using both the normal and the heavy isotope. He used an interference method, employing the Fabry-Perot etalon and took great care to work with a discharge tube at low pressure and very low temperature. Results were good.[11] The two main peaks shown in Fig. 6 are separated by about one-fiftieth of the wave-length separation between the familiar doublet yellow lines of sodium. The small peak just to the left of the second major line is the most interesting line. It is mainly due to the transition between the $^2P_{\frac{1}{2}}$ state, third radial quantum level, and the $^2S_{\frac{1}{2}}$ state, second radial level. Its close, strong neighbor is the corresponding transition between excited states, but $^2D_{3/2}$ to $^2P_{\frac{1}{2}}$. The spacing between these lines is given by the Dirac theory, and is due essentially to relativistic effects. The separation of the main components can also be computed. Both of these values are off, not by much, but by what seems an amount beyond the error of the experiment. The small peak, for example, is too far from its strong neighbor by about its own width. Were the Dirac theory exact, the shape of the line would be much different; the little peak would be merely a smooth step, and not a distinct peak on the larger line. It was suggested at that time, nearly ten years ago, that the discrepancy was due to a shift upward by a small amount in the position of the S state involved.[12] Both the difficulties noticed here, and a number of other irregularities of fine-structure doublets, would have been resolved by this assumption. But the Dirac theory was a good theory, and the discrepancy observed, while disturbing, was small enough to shrug off. Other experiments confirmed it only partially, and interfering lines from the molecule were blamed for the disagreement. The facts were clear, but they were somewhat uncomfortably minimized. This discrepancy is after all in the last decimal place of the old spectroscopy. In 1947, a decade later, the hydrogen atom was examined by the new spectroscopy. This is based on the absorption of quanta in the radiofrequency region, and often involves identifying atomic states by observing the atoms in motion in an atomic beam under the action of impressed fields. Two workers of the Columbia group, where under Professor I. I. Rabi this beautiful technique had flowered,[13] have studied the hydrogen atom in an experiment of enormous ingenuity and precision. The scheme of Lamb and Retherford was the following.[14]

Instead of the H_α line components, all of which involve the transition of an electron between two different radial states, they would observe the very much lower energy transitions in which transitions take place between two atomic states which differ only in the angular momentum properties, not at all in the radial motion. They thus measure the fine structure directly, and not as a difference of two Balmer transitions. Such states lie very close on an energy scale: the quantum energy must be measured in microvolts instead of in the electron volts of the optical region. The wave-lengths are not those of red light, but rather some three centimeters. Here any old radar man would feel shame if he could not measure frequency to within a few percent! Molecular hydrogen, largely dissociated by heating, is allowed to emerge from an oven into an atomic beam apparatus. Across the beam of atoms an electron stream passes, exciting occasional atoms—about one in a hundred million—to the second radial state, with angular momentum given in the usual way as $^2S_{\frac{1}{2}}$. Radiation from this state to the ground state is very slow, and the atoms move in the beam in this metastable state until they strike a metal target. There such excited atoms can be recognized by a curious process, long used in the molecular-beam work to detect some special beams. They give their excess energy to metallic electrons, falling themselves to the ground state. The electrons leave the metal as though by thermal emission, producing a secondary electron current, measured with an electrometer circuit. If now any process causes the atoms in flight to make a transition to any other state which is not metastable—say one of the 2^2P levels—the atoms will quickly decay to

[11] R. C. Williams, Phys. Rev. **54**, 561 (1938).
[12] S. Pasternack, Phys. Rev. **54**, 1113 (1938).
[13] See J. M. B. Kellogg and S. Millman, Rev. Mod. Phys. **18**, 323 (1946).
[14] W. E. Lamb and R. C. Retherford, Phys. Rev. **72**, 241 (1947).

companies the disintegration of the Bristol heavy mesons. Such an identification is unsure, but it does lead to a rather neat theory.

Even this is not all. More cloud-chamber pictures, this time from Manchester,[8] and some mass measurements from Berkeley cosmic-ray workers,[9] and a whole series of rather indirect experiments done in the Caucasus by Russian cosmic-ray workers,[10] all point to more mesons, perhaps about a thousand times the electron mass, both charged and neutral. None of these pieces of evidence is conclusive; taken together they are at least disquieting.

The study of nuclear forces, the best use of high energy machines, the superenergy cosmic-ray phenomena, especially the giant showers—these are the problems before the workers in mesons. 1947 brought the simplest and most unsatisfactory account of mesons to a close, and entered a new era of meson sophistication. There is wide opportunity for many approaches to a partial solution of many problems. It is sure that the meson exists; and now we have a whole family of these intermediate particles, of positive, negative, and zero charge. The final unraveling of this complex and the description of nuclear forces with its aid is surely a task for a decade, even at the present rate. More mesons are not to be excluded; the unfundamental character of our fundamental particles now cries for explanation. It is certain that the reviews of the years to come will continue to illuminate this new and exciting landscape.

SPECTROSCOPY, OLD AND NEW

For a long time, now nearly thirty-five years, the atomic spectra, and especially the spectrum of the simplest atom, hydrogen, have been the touchstones of physical understanding. The Bohr theory of the atom won its success from the Balmer series; the Sommerfeld relativistic modification, from the doublet structure; and the Dirac theory, from its complete agreement with the correct Sommerfeld doublet positions, and from the intensities which it correctly gave. From the side of theory, the system was the simplest we know; from the side of experiment, the technique was enormously precise and refined. It was the place for the test of fundamental ideas.

Not all was well. In 1937 (see Fig. 6) R. C.

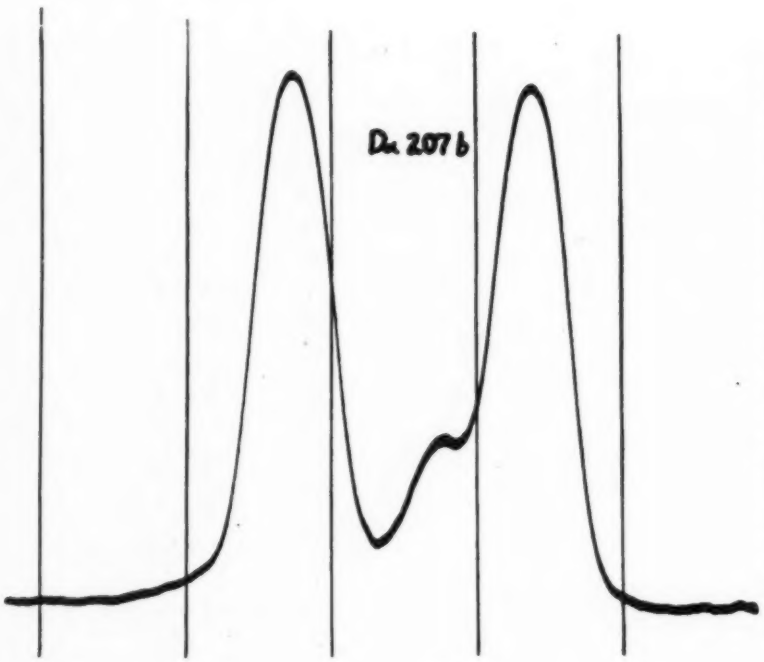


FIG. 6. This is a microphotometer tracing of a high resolution spectrograph of the strong red line of the hydrogen atom. (This is actually a photo of the D_α line in deuterium, hydrogen of weight two.) The two strong peaks are the lines of the main transitions, and the small satellite on the right-hand line is the transition from the state $3^2P_{\frac{1}{2}}$ to the state $2^2S_{\frac{1}{2}}$. According to the Dirac theory, the small peak should lie about 0.03 cm^{-1} closer to its strong neighbor than it does in this tracing. The vertical lines on the tracing are spaced about one-fifth cm^{-1}. While the line is not well resolved, it seems clear that moving it that far to the right would merge it nearly completely with the righthand doublet line. The line shift so measured was given the correct interpretation in 1938. The tracing was kindly supplied by Professor R. C. Williams; it appears in The Physical Review **54**, 561 (1938). In the same year S. Pasternack correctly evaluated the shift of the S level of hydrogen, which would explain this and other small spectroscopic effects. It took ten years to come to believe this result.

[8] Rochester *et al.*, private communication.
[9] J. G. Retallack, Phys. Rev. **72**, 742 (A) (1947).
[10] Alichanian, Alichanov, and Weissenberg, J. Phys. U.S.S.R. **11**, 97 (1947).

Williams, then at Cornell, made the best of a number of precision and high resolution spectroscopic examinations of the fine structure of the prominent deep red line of the hydrogen atom—the well-known H_α line—using both the normal and the heavy isotope. He used an interference method, employing the Fabry-Perot etalon and took great care to work with a discharge tube at low pressure and very low temperature. Results were good.[11] The two main peaks shown in Fig. 6 are separated by about one-fiftieth of the wave-length separation between the familiar doublet yellow lines of sodium. The small peak just to the left of the second major line is the most interesting line. It is mainly due to the transition between the $^2P_{\frac{1}{2}}$ state, third radial quantum level, and the $^2S_{\frac{1}{2}}$ state, second radial level. Its close, strong neighbor is the corresponding transition between excited states, but $^2D_{3/2}$ to $^2P_{\frac{1}{2}}$. The spacing between these lines is given by the Dirac theory, and is due essentially to relativistic effects. The separation of the main components can also be computed. Both of these values are off, not by much, but by what seems an amount beyond the error of the experiment. The small peak, for example, is too far from its strong neighbor by about its own width. Were the Dirac theory exact, the shape of the line would be much different; the little peak would be merely a smooth step, and not a distinct peak on the larger line. It was suggested at that time, nearly ten years ago, that the discrepancy was due to a shift upward by a small amount in the position of the S state involved.[12] Both the difficulties noticed here, and a number of other irregularities of fine-structure doublets, would have been resolved by this assumption. But the Dirac theory was a good theory, and the discrepancy observed, while disturbing, was small enough to shrug off. Other experiments confirmed it only partially, and interfering lines from the molecule were blamed for the disagreement. The facts were clear, but they were somewhat uncomfortably minimized. This discrepancy is after all in the last decimal place of the old spectroscopy. In 1947, a decade later, the hydrogen atom was examined by the new spectroscopy. This is based on the absorption of quanta in the radiofrequency region, and often involves identifying atomic states by observing the atoms in motion in an atomic beam under the action of impressed fields. Two workers of the Columbia group, where under Professor I. I. Rabi this beautiful technique had flowered,[13] have studied the hydrogen atom in an experiment of enormous ingenuity and precision. The scheme of Lamb and Retherford was the following.[14]

Instead of the H_α line components, all of which involve the transition of an electron between two different radial states, they would observe the very much lower energy transitions in which transitions take place between two atomic states which differ only in the angular momentum properties, not at all in the radial motion. They thus measure the fine structure directly, and not as a difference of two Balmer transitions. Such states lie very close on an energy scale: the quantum energy must be measured in microvolts instead of in the electron volts of the optical region. The wave-lengths are not those of red light, but rather some three centimeters. Here any old radar man would feel shame if he could not measure frequency to within a few percent! Molecular hydrogen, largely dissociated by heating, is allowed to emerge from an oven into an atomic beam apparatus. Across the beam of atoms an electron stream passes, exciting occasional atoms—about one in a hundred million—to the second radial state, with angular momentum given in the usual way as $^2S_{\frac{1}{2}}$. Radiation from this state to the ground state is very slow, and the atoms move in the beam in this metastable state until they strike a metal target. There such excited atoms can be recognized by a curious process, long used in the molecular-beam work to detect some special beams. They give their excess energy to metallic electrons, falling themselves to the ground state. The electrons leave the metal as though by thermal emission, producing a secondary electron current, measured with an electrometer circuit. If now any process causes the atoms in flight to make a transition to any other state which is not metastable—say one of the 2^2P levels—the atoms will quickly decay to

[11] R. C. Williams, Phys. Rev. **54**, 561 (1938).
[12] S. Pasternack, Phys. Rev. **54**, 1113 (1938).
[13] See J. M. B. Kellogg and S. Millman, Rev. Mod. Phys. **18**, 323 (1946).
[14] W. E. Lamb and R. C. Retherford, Phys. Rev. **72**, 241 (1947).

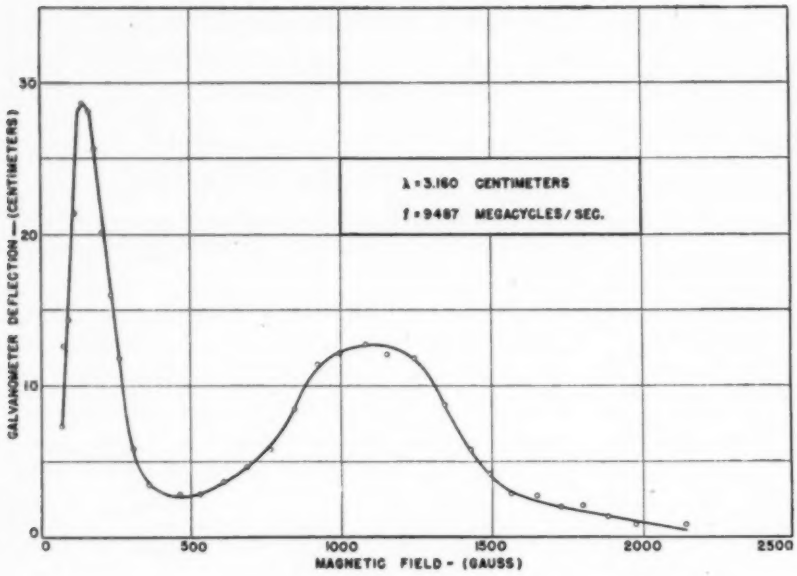


FIG. 7. This is the new spectroscopy. The curve plotted shows the change of current of metastable atoms resulting from the absorption of microwave quanta, plotted as a function of magnetic field As explained in the text, this really shows the Zeeman effect of the lower level of the fine-structure transitions shown in William's optical spectrograph. It represents an enormous magnification of the frequency spectrum: three-centimeter waves replace red light The width of the lines shown here is in part real, in part due to geometrical and instrumental effects. The curve was kindly furnished by Professor W. Lamb at Columbia.

the ground state by emitting one of the lines of the Lyman series. The atoms then arrive at the target in their ground state, and are not detected. One scheme which would induce such transitions might be the absorption of microwave quanta, which would induce transitions from the metastable S level to one of the corresponding P states. If now the current of metastable atoms is plotted as a function of microwave frequency, a resonance curve will appear, minimum beam current showing the resonant frequency of the transition. In Fig. 7 there is such a curve. Actually, a somewhat more complicated effect is used for this curve. The atoms are allowed to enter a magnetic field, and the fine structure states split by the anomalous Zeeman effect. Then the curve is run as a function of the applied magnetic field at a fixed frequency. The curve shows the change in galvanometer deflection caused by the microwave radiation. The two peaks are two of the four lines of the Zeeman effect for the transition between the states of different magnetic quantum number in the transition $2^2S_{\frac{1}{2}}-2^2P_{\frac{1}{2}}$. Without magnetic field, the resonance would come at about 9950 megacycles, and would be single. The magnetic field values for which resonance occurs at the frequency applied are now easily measured, and lead to a most precise measurement of the hydrogen fine structure. It should be observed that the splitting between the two right-hand peaks in Fig. 6, the H_α photograph, is represented in the Lamb-Retherford work by the measured microwave frequency of resonance, *not* by the Zeeman splitting between the two peaks. The Zeeman splitting is not more than about a tenth of the main effect, even at 1000-gauss field. But even this splitting is well measured.

The Lamb-Retherford result cannot be doubted. The two levels $2^2S_{\frac{1}{2}}$ and $2^2P_{\frac{1}{2}}$, which were supposed on the Dirac theory to lie at exactly the same energy, are in fact split; the $2^2S_{\frac{1}{2}}$ state lies higher by about 0.033 cm^{-1}, just one-eleventh of the predicted separation of the two P states, of $\frac{1}{2}$ and $\frac{3}{2}$ total angular momentum. The P separation seems to agree within present experimental error with the Dirac theory. Just this result was suggested ten years ago on the basis of the optical spectra, but theorists could avoid coming to grips with the problem until the Lamb-Retherford result showed unambiguously and precisely just how far the Dirac atom was wrongly built.

The full explanation seems already to be at hand. Even a decade ago the suggestion had been made of how to correct the Dirac theory. But this year theorists found it necessary to answer the question, and they found how to do it. It will not be possible to give here the details of the theory now reaching formal completion, but it will be our attempt to understand the nature of the problem and its solution.

The Dirac theory of the atom treats a simple system, the electron moving in the field of a point

positive charge. It is clear that this is an abstraction from reality. For example, the nucleus is not a point. It is a proton with a definite size and a magnetic moment, for example, which cannot be predicted by any theory today. But the effect of such an idealization is not what the measurements show. Clearest proof is that the deuterium atom shows a very similar level shift. This is convincing for the idea that it is an atomic and not a nuclear problem we here face. Optical spectral measurements[15] done in 1947 in order to test the theoretical explanation show split components of lines of the spectrum of ionized helium which fully confirm the general ideas, and even measurements on nuclear magnetic moments and hyperfine structure—a very different field—lead to a similar interpretation. So the same idea has found precise expression in more than one experiment.

What is the idealization of the Dirac atom which we now are forced to re-examine? It is that the electron moves about the proton but does not radiate, and has no connection with the electromagnetic field except as it responds to the charge of the proton. It is as though one were to calculate the normal frequencies for vibration of a perfect piano string without taking account of the fact that it can radiate sound. That is a very good approximation, and the frequencies so found are very nearly those observed. But actually the piano string is not kept in a vacuum, but interacts with the air of the room. Two effects will be noticed because of the air of the room: first, the motion of the string, once started, will not persist forever, because energy will be lost by the radiation of sound. This will introduce a damping factor, the familiar exponential decay of the amplitude with time. This modulation on the otherwise perfect sine wave produced by the string has the effect of introducing side-bands; the single frequency will now not be perfectly sharply defined, but a small band of frequencies will be emitted by the string, even if it vibrates in the fundamental mode of its air-free motion. Such an effect will, of course, be the greater the greater the radiation; for a string which vibrates in a dense atmosphere, the frequencies might be considerably spread. This is a good analog for the atom. The spectral lines are not perfectly sharp; they have a characteristic natural width which depends on the rate at which the atom can radiate quanta, and so leave its excited states. In quantum mechanics only the ground state is completely sharp. So much is quite familiar; everyone knows of the natural radiative width of a spectral line. It is small because radiation is a relatively unimportant effect. The atomic "piano string" is little damped by the "atmosphere" into which it radiates. The Dirac approximation, which is to treat the level as quite sharp in the original problem and then to correct by adding the broadening effect, is satisfactorily accurate.

But there is another property of the string-air system. The presence of the room air will act to increase the effective mass of the string; it must move the air nearby when it vibrates. This will not only broaden the sharp frequency, but will also slightly shift the fundamental frequencies of the string from the value they would have had in a perfect vacuum, without sound radiation. This is just the Lamb-Retherford shift, as Bethe has first shown.[16] It corresponds to an effective change in the characteristic atomic energy levels, related to the line broadening in the same way as the shift of the piano string frequency is related to its damping. The damping is a resistive effect; the level shift is in the same electrical language a reactive effect.

Why was this hard to discover? The answer is simply this: the energy of an electron in the Bohr orbit is the sum of its rest mass plus its energy resulting from the atomic state. If you now calculate—and this is a problem as old as Lorentz—the energy of an electron in interaction with the electromagnetic field in free space—not in an atom—you get an infinite, and hence a senseless result. This follows whenever you refrain from giving the electron a structure, a definite size. And no clear-cut way of giving it a size has been found, though it is a problem a generation old. The great advance of 1947 was not so much finding a logical way to make the infinite result come out calculable, as finding a scheme by which the free electron with its infinite electromagnetic mass could be compared with the still infinite electromagnetic energy of the electron bound in hydrogen. These two showed a difference, *and that difference is just the Lamb shift.* The infinities are false and will disappear from a future theory;

[15] J. E. Mack and N. Austern, Phys. Rev. **72**, 972 (1947).

[16] H. A. Bethe, Phys. Rev. **72**, 339 (1947).

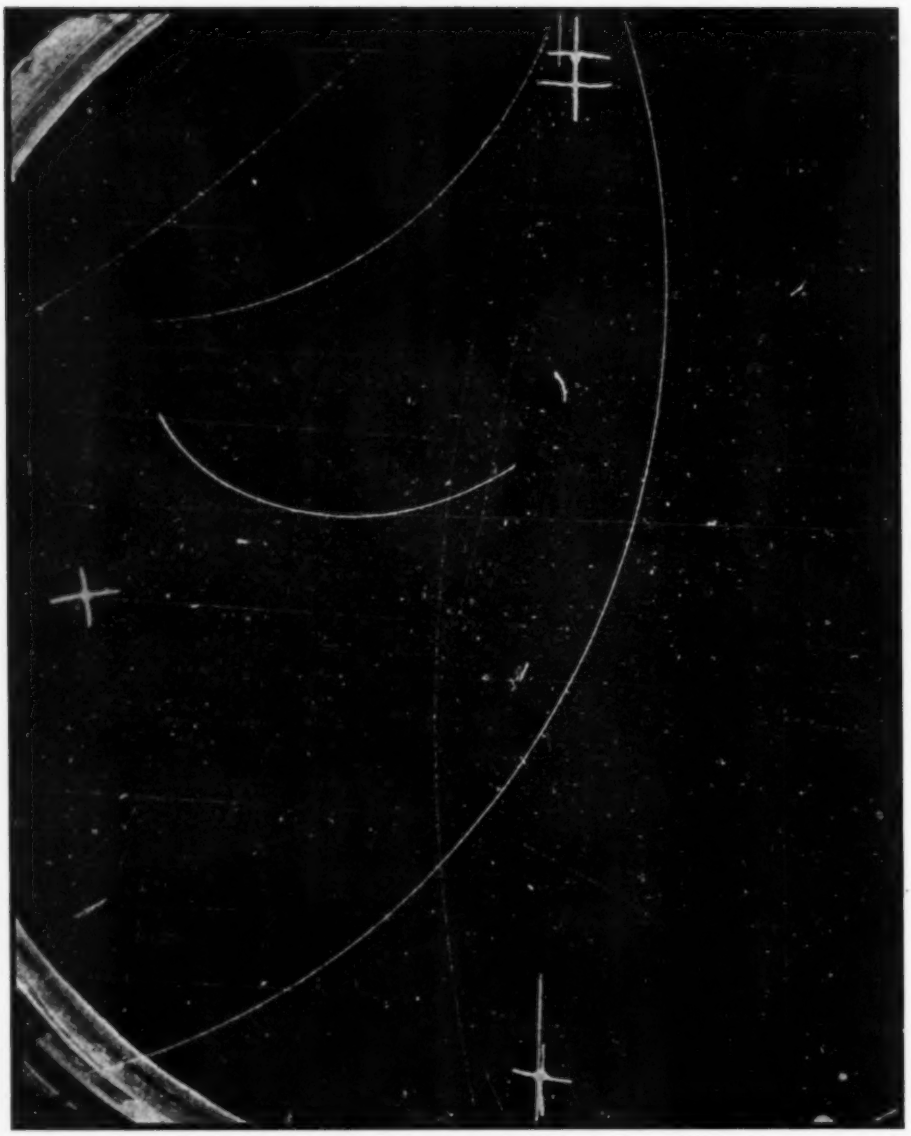


FIG. 8. The cloud chamber at Berkeley, with a strong magnetic field. The two heavy tracks are a low energy, strongly curved proton track and a straighter high energy proton. These protons are recoils from a hydrogen containing gas placed in the cloud chamber. It is very likely that the high energy knock-on proton is in fact just the incident neutron, now become a proton by the action of the exchange forces, as described in the text.

the difference is real and can be computed from today's concepts.

Welton has pointed out a very simple way of regarding the effect.[17] One consequence of the Heisenberg uncertainty principle is that in quantum mechanics the number of quanta in the electromagnetic field can be fixed precisely only at the price of admitting fluctuations in the electric and magnetic fields in space. In this way we may say that an atom shows the photo-effect because the electric field, which on the average was only very small, corresponding to a few quanta being present, yet can transiently be strong enough to eject an electron from its orbit. Even in a region of space where there are no quanta, then the electric and magnetic fields, on the average certainly zero, fluctuate widely. These effects are known as the vacuum fluctuations of the electromagnetic field. They are, of course, a wholly quantum-theoretic result. How will an electron respond to these fluctuating fields? It will, of course, move in them, wildly moving about in a kind of Brownian motion in free space, colliding with the quanta which are only transiently present. A free electron will respond; so

[17] Welton, Phys. Rev. **73** (1948), in press.

will a bound electron. If you allow the possibility of response to transient quanta of all frequencies, you get the infinite energy of the free electron. But forget this inconsistency, and compare a free electron with a bound one. Quantum by quantum, the effect will be the same. But the bound electron is forced by its "Brownian motion" to tremble in the potential field which binds it as it executes its orbital motion. Its average energy then will be a little changed. This is the Lamb level shift. It is not hard to see that the effect will depend on the gradient of the force field in which the electron moves, or the second derivative of the potential. For the fluctuating motion is random in direction, and if the force were constant as much work would be done by the electron on the binding field as by field on electron, and no change would be noticed. Only the lack of cancellation will show up; this depends on the second derivative of the potential. Since the motions are to be thought of as very small ones, higher derivatives will have little effect. This argument can be made precise, and a good value gotten for the level shift.

This is not the whole story. For the vacuum is peopled not only with the fluctuating electromagnetic fields, but with the unnoticed electrons in negative energy states introduced by Dirac. It is these electrons which can be promoted in energy. When this is done the electron is now a visible ordinary one, and the "hole" it left behind behaves like a positron, the "donkey" electron which moves backward when you pull it acting like a positive charge. In the vacuum, where no electrons or positrons are present on the average, there is a fluctuating charge density corresponding to the transient appearance of positrons and electrons, promptly annihilating each other. With these charges and currents the electron can also interact, and this somewhat modifies the level shift. This latter effect also contributes to an

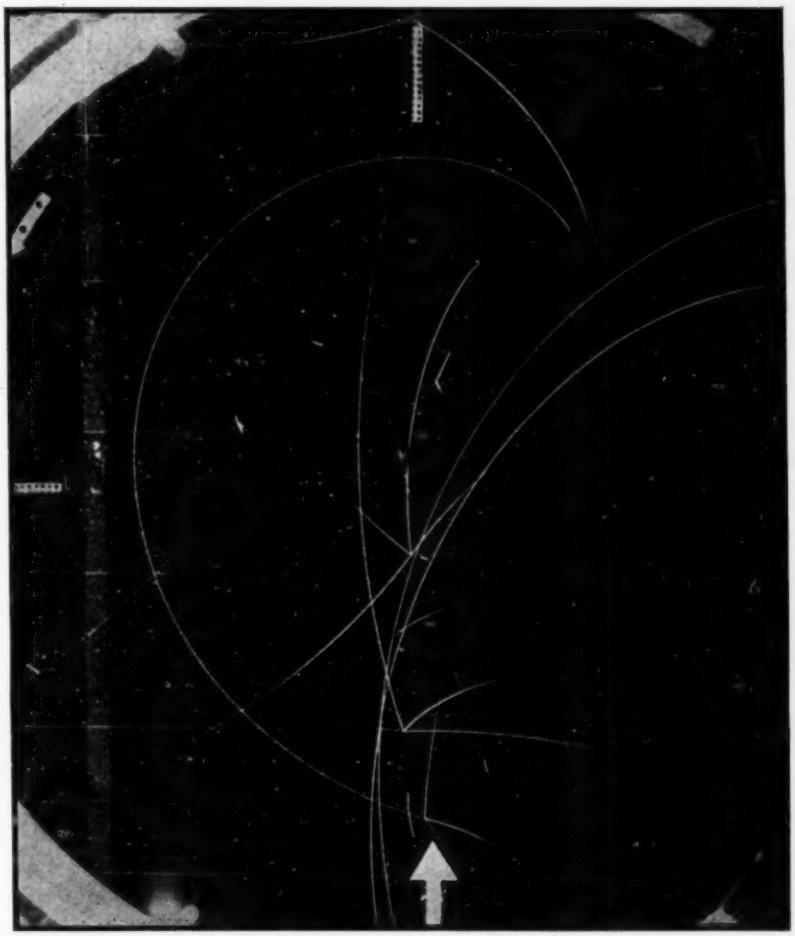


FIG. 9. This is a photograph of a cloud chamber with a strong magnetic field, taken at Berkeley with the 184-in. f-m cyclotron. The 90-Mev neutrons are incident on the nuclei of the cloud-chamber gas, and the tracks shown include two star processes, where several heavy charged particles radiate from a single point. These are artifical equivalents to the cosmic-ray stars shown in Figs. 4 and 5, but with no mesons involved. Neutrons cause these star processes.

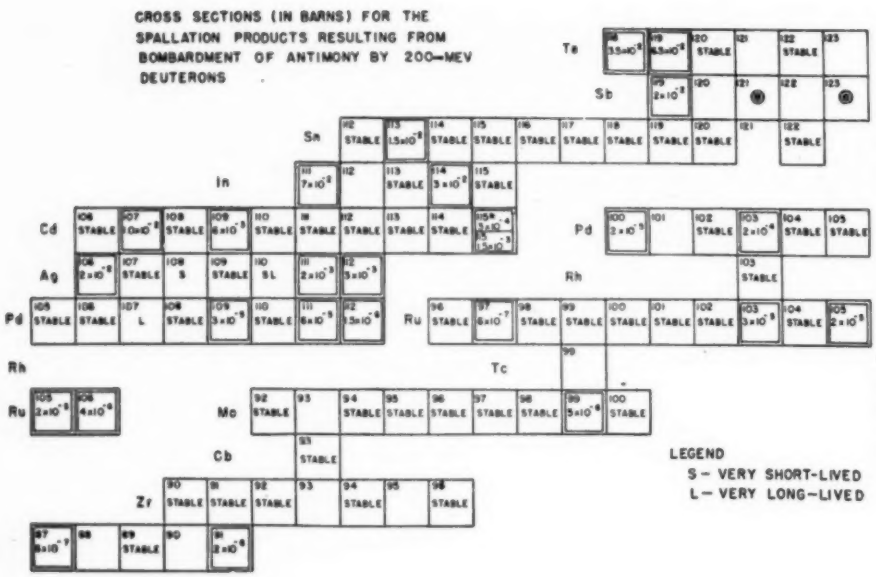

CROSS SECTIONS (IN BARNS) FOR THE SPALLATION PRODUCTS RESULTING FROM BOMBARDMENT OF ANTIMONY BY 200-MEV DEUTERONS
LEGEND
S - VERY SHORT-LIVED
L - VERY LONG-LIVED

FIG. 10. This is radiochemistry at high energies. The sample of antimony, containing the two target isotopes (so marked) 121 and 123, yields on bombardment with 200-Mev deuterons all the activities marked with a square in the isotope chart. Most frequent are the collisions in which a couple of nucleons are knocked out without much heating of the nuclear matter, but rarely many particles, up to twenty-odd, can be boiled off. The chart was furnished by Dr. Perlman of Berkeley.

effect not correctly calculable from the "Brownian motion" idea alone, a small deviation from the Dirac value of the magnetic moment of the electron. This has also been confirmed[18] experimentally in 1947 by comparison of calculated and measured Zeeman splittings, done with great accuracy by atomic beam magnetic resonance effects in gallium and sodium spectra. This shows a distinctly larger spin magnetic moment for the atomic electron than exactly the unit Bohr magnetron of theory, if we assume the orbital effect unchanged. Schwinger[19] and a series of investigators have by now developed the theory to a point where all these small effects can be essentially predicted theoretically. It is to be pointed out that while no very new concepts have yet come in, this formally beautiful work is of the greatest importance. The effect is small for electrodynamics, and the really new corrections to the quantum analog to Maxwell's equations are hard to find. None of the experiments yet foreseen require that the still not quite clear way to make the infinite terms finite be given correctly. But in meson theory the coupling of the source with its meson radiation is large, not weak as for ordinary quanta. Here we may expect that the new theoretical approach will meet a stringent test. Can it predict the reaction of the meson field? This determines not last-decimal-place quantities, but major ones like the magnetic moments of the nuclear particles. It is too soon to say, but the year's progress in theory not only goes far towards a satisfying and spectacular clearing-up of old difficulties, not only confirms very precise up-to-date spectroscopy, but presages well for the treatment of the nuclear forces themselves. We do not yet have the theory of the future, but its framework is well in place.

THE HIGH ENERGY MACHINES

There are two dozen or so laboratories in the world busy on the construction of machines to make particles or radiation above 100-Mev energy. Not one of these was completed in 1947. Only the 100-Mev betatron at General Electric in Schenectady and the 200-Mev frequency-modulated cyclotron at Berkeley are now working in this range. Both these machines were operating in 1946, and the betatron even before.

In last year's review the hope for artificial mesons was high. The first section this year gives an account of how that hope has waned, and waxed again. Still no 300-Mev machine is working; still, no mesons. Under construction at Birmingham in England is a 1300-Mev proton synchrotron, and designs at least have been made at Berkeley and at Brookhaven for a machine at eight or ten billion electron volts.

[18] P. Kusch and H. M. Foley, Phys. Rev. **72**, 1256 (1947); Nafe, Nelson, and Rabi, Phys. Rev. **71**, 914 (1947).
[19] J. S. Schwinger, Phys. Rev. **73**, 415 (1948).

But much has been learned even without mesons. Physics above 100 Mev is pretty much physics at Berkeley, and Figs. 8, 9, and 10 will give some idea of what they are doing. There are several important results.

1. *The exchange nature of nuclear forces.*—It is more than ten years since the idea of exchange forces for the nuclear particles gained pretty general acceptance. They are thought of as simple analogies to the kind of forces which unite two protons and an electron into a hydrogen molecular ion, for example. This kind of chemical bond depends upon the possibility of the bonding electron being found near either proton, i.e., of its "exchanging" its position from one proton to the other. Such a chance for exchange brings in a purely quantum-mechanical effect, which makes the energy of the two atoms much lower when close enough so that the electron exchange is more readily possible than when the atoms are far apart, and the electron cannot exchange positions. The important consequence of such a bond is that the approach of a third proton will not result in a lowered energy. The additional chance for exchange gives no further binding. Even more striking is the two-electron kind of bond in the neutral molecule of hydrogen. Here the two electrons must have opposite spins to lead to binding. A third atom must present the same electron spin as at least one of the two already present, and will in fact be repelled. Such forces are said to be saturated; the number of effective bonds is limited. The chemists know such forces well.

When these forces were postulated for nuclei, they were regarded as simply formal assumptions, needed to give "saturation." It is a well established fact that in heavy nuclei the nuclear particles appear to interact strongly with only a few other nucleons; the binding energy *per particle* is about the same for nuclei of all weights. So exchange forces were needed.[20]

With the advent of meson theory a more physical picture could be given for such forces. They involved the exchange of mesons between nuclei, just as chemical forces involve the exchange of electrons between atoms.

Recent experiments at Berkeley have involved the scattering of the high energy neutrons—about 90 Mev—from protons. The results were somewhat striking. A large fraction of the protons which were struck appeared to go forward, in about the direction of the neutron beam, with nearly the whole of the original neutron energy. This could be explained if the strength of the nuclear forces was very great, so that these head-on collisions could occur quite frequently. But we know that the nuclear forces are not that strong; the low energy scattering and the binding energies of light nuclei argue for a nuclear potential of only about 10 Mev. Another process is taking place, one predicted fifteen years ago. The collision of neutron with proton often results, even when the collision is not a direct head-on encounter, in the exchange of place between neutron and proton. The incoming neutron is not much deviated by the collision, but it becomes a proton *en passant*, and the target nucleon, now a neutron, recoils with a modest energy. This is the straightforward demonstration of the reality of exchange forces. How much such forces contribute to the total interaction can also be answered by precise measurements of the same sort. The cloud-chamber photographs of Fig. 8 show such knock-on protons,[21] some of which may be the very neutron sent in, now transformed by the exchange force.

2. *Nuclear reactions at high energy.*—For ordinary nuclear reactions, with particles having some 10,000,000-volts energy, the nucleus has been compared with good results to a liquid drop. In this the hundred-odd nucleons of a middle-sized nucleus constantly jostle each other. A particle striking the drop "warms it up," imparting the incident energy to a few nucleons which quickly pass it on to their neighbors, so that the whole drop is heated. The drop may cool by evaporation of some of its particles or by radiating gamma-rays. Such a picture really works in accounting for nuclear reactions of the familiar sort.

But 100-Mev neutrons have led to a different picture. They move so fast that they cross the whole nucleus before the other particles have moved even a little. Moreover, the fast particle has a rather small cross section for collision with a

[20] H. A. Bethe, *Elementary Nuclear Theory* (John Wiley and Sons, Inc., New York, 1947), Chapter XIV.

[21] W. Powell *et al.* (in press).

single nuclear particle. The nucleus no longer acts like a liquid drop, but rather like the snapshot of a small bubble of gas. The fast particle may go right through a middle-sized nucleus without touching any particles; the nucleus becomes partly transparent. But if a particle does get struck, things can indeed happen. If the struck particle lies on the surface of the nucleus, and takes perhaps 20 Mev from the incident projectile, it will leave without more ado, and a rather simple nuclear reaction will have taken place. But it may be that the struck particle lies deep within the nucleus. It has a high energy, but not at all enough to cross the nucleus on its way out without striking other nuclear particles. Thus it will share its energy with the other particles, and the nucleus is again heated. The nucleus may now be very hot indeed, and boil off a large number of particles before reaching its ground state. In this way then, high energy reactions are extremely copious and varied. The comfortable days of nuclear chemistry, when you knew that only two or three reaction products were at all likely, are ended at high energy. Two different experimental approaches are shown in accompanying figures. Figure 10 shows the rich and complex products of the bombardment of the two isotopes of antimony with 200-Mev deuterons.[22] In Fig. 9 the center of the photo shows several tracks radiating from a single point. These are the product particles from some rather complex reaction of the fast neutrons with the light atoms of the cloud-chamber gas. Such figures in cloud chamber or photographic plate are familiar from cosmic-ray work, where they are called "stars."[23] The study of stars and the radio-chemistry of high energy reactions provides a whole new field of nuclear physics.

This work, and the verification of the idea of the transparent nucleus described above,[24] by cross-section measurements[25] and by study of the mechanism of the deuteron break-up at the beryllium target in Berkeley[26] to make neutrons, form the first successes of nuclear physics at high energy.

[22] Seaborg, Cunningham, Hopkins, Lindner, Miller, O'Connor, Perlman, and Thompson, Phys. Rev. **72**, 740 (A) (1947).
[23] W. Powell, Phys. Rev. **72**, 739 (A) (1947).
[24] R. Serber, Phys. Rev. **72**, 1114 (1947).
[25] Helmholtz, McMillan, and Sewell, Phys. Rev. **72**, 1003 (1947).
[26] Cook, McMillan, Peterson, and Sewell, Phys. Rev. **72**, 1264 (1947).

NUCLEAR PHYSICS: TRADITIONAL ENERGIES

The year saw a by no means diminished activity in the more familiar domain of nuclear physics at ordinary energies. Here the old lines of effort continue, and there is an increasing tendency to do experiments precisely and well. A good measurement of most nuclear quantities was once one which had a probable error of twenty-five percent. Today experience and the many new techniques make possible some work to better than a percent, and, of course, the r-f spectroscopy gives its usual high precision.

The neatest experiment of the year was one done with the chain reactor at the Argonne Laboratory in Chicago by Professors Fermi and Marshall.[27] It was an attempt to observe any specific interaction between the neutron and the electron. Since both have magnetic moments, there is an obvious magnetic interaction. This was not observed, for the experiment was done with an atom, actually xenon, in which there is complete cancellation of the electron magnetic moments. What was looked for was some new, specific interaction. Perhaps the neutron is, in fact, a nuclear core surrounded by a cloud of mesons, so that some of the time the system is like a proton accompanied at distances of 10^{-13} cm or so by a negative meson. This would have a very small interaction with the electron whenever the electron got close to the center of the nuclear particle.

The method used was actually suggested by E. U. Condon in 1936.[28] It depends on the fact that for any scattering process the scattering will be spherically symmetrical provided that the wave-length of the scattered wave is large compared to the size of the scattering center. This is well known in acoustics, in radio, and indeed in quantum mechanics. Large wave-lengths mean pure diffraction, and the opposite of specular reflection, which is isotropic scattering. If thermal neutrons with wave-lengths of several angstroms are scattered from nuclei alone, the scattering will be spherically symmetric. But if the electrons around the nuclei take part, there will be a devia-

[27] E. Fermi and L. Marshall, Phys. Rev. **72**, 1139 (1947).
[28] E. U. Condon, Phys. Rev. **49**, 459 (1936).

tion from isotropic scattering, for the atomic structure is not small compared to the neutron wave-length. The scatterer used was xenon gas. Such a choice was required if all magnetic effects were to be avoided, a non-paramagnetic atom had to be used. If all interference of waves coming from different atoms is to be avoided, a gas must be used. And the rather high cross section of xenon made the effect sizable, while its heaviness reduced the corrections resulting from the thermal motion of the atoms.

A carefully collimated beam of thermal neutrons was allowed to enter a tank containing xenon at about one atmosphere. Cadmium diaphragms allowed the neutron beam to see only the gas or the Cd lining of the tank. The scattered neutrons were counted at a forty-five degree angle by the ordinary BF_3 neutron counters. A scheme for turning the whole scattering apparatus end-for-end was used to minimize geometrical asymmetries of construction. After corrections the fractional difference between 45° and 35° scattering turned out to be almost zero, actually 0.0005 ± 0.0085. This result gives not the ratio of the cross section for the electrons compared to the nucleus, but the ratio of amplitude of scattered wave. Moreover, many electrons are involved. The final result is that the strength of interaction between electron and neutron is about one in fifty thousand as great as that between neutron and xenon nucleus. This result is remarkable in that it corresponds to the measurement by this diffraction method of an effect which, if acting alone, would lead to a cross section between neutron and electron of only some 10^{-34} cm^2. No clear-cut theoretical result is available on meson theory for the value of the interaction. The error in this experiment is about the size of the roughly estimated theoretical values. The measured null effect, since it is much smaller than the error, would if taken at face value be hard to reconcile with meson theory expectations. But a precision ten times higher would be needed to lead to a real contradiction. It is now sure, however, that no interaction much larger than that of the meson theories does exist between neutron and electron.

A rather similar but less convincing experiment, using molten lead, was done by the Columbia group of Rabi, Havens, and Rainwater. Their result is in essential agreement with that of Fermi.[29]

Another slow neutron experiment of interest was reported in 1947. It was the establishment of the resonances in neutron scattering.[30] These sharp peaks in the neutron transmission curve have long been studied. They were known to come from a high probability of absorption of slow neutrons whose energy coincided with that of a quantum state in the compound nucleus. The theory had long pointed out that such levels should show not only strong absorption but also strong scattering. This was verified for the first time unambiguously at the Argonne Laboratory. The experiment was simple: an annular counter surrounded a cobalt foil. The counter could detect neutrons only if they were scattered by the enclosed foil. This foil was then subjected to a beam of neutrons which had been passed through another cobalt foil. The detected neutrons would, ideally, be only those which had been strongly absorbed by the first cobalt foil. Its cross section for absorption would appear much higher than if all neutrons passed by the foil, even those far from resonance energy, were scattered and detected. Such an effect is indeed observed. This work is reminiscent of the old neutron days, when thermal neutrons were used for such experiments. It is only the intense sources and long experience of today which enables such work to go in the resonance region above 100-ev neutron energy. This work is less spectacular than most we have reported, but it deserves mention as the representative of a large number of results, both in experiment and theory, which are finally taking the guesswork and the confusion out of the behavior of nuclei in the region of their ground states and a few million volts above. It is as such a representative that I have elected to report it. The literature is filled with many more examples of the maturing of nuclear physics.

To mark a different point, there is the very recent work of the beta-ray spectrograph workers at Indiana.[31] They have paid special attention to

[29] Havens, Rabi, and Rainwater, Phys. Rev. **72**, 634 (1947).
[30] N. H. Barbre and M. Goldhaber, Phys. Rev. **71**, 141 (A) (1947); Harris, Langsdorf, and Seidl, Phys. Rev. **72**, 866 (1947).
[31] L. M. Langer and C. S. Cook, Phys. Rev. (in press).

the low energy end of the energy distribution of the beta-particles from Cu^{64}. They are especially concerned with electrons and positrons—this substance emits both—below 100-kv energy. This is a difficult research; such electrons are heavily absorbed and scattered by quite thin pieces of material, and great care is required to ensure against results falsified by such absorption or scattering. They have done the job, and they confirm and extend the results obtained earlier. There must be hundreds of experiments which examine carefully the upper end of the beta-spectrum; only these two looked carefully at low energies. Their results agree: there are very few electrons at low energy, and too many positrons. Theory has been able to give a fairly convincing account of all the phenomena of beta-decay, but there is no explanation yet of this absence of the low energy electrons. Every scheme which has been thought of to modify the theory either is inadequate to explain the really striking size of the effect, or else throws away most of the worth while old results of beta-theory calculations. The problem is all the more worrisome because it involves nuclei and electrons in a region where we thought we understood what happened, a region not very different from that of the x-ray levels of heavy atoms. Much more experiment is needed. Perhaps this is some special effect in this particular nucleus. It is to be hoped that next year will see those who measure beta-ray spectra paying as much attention to the low energies as to the high ones.

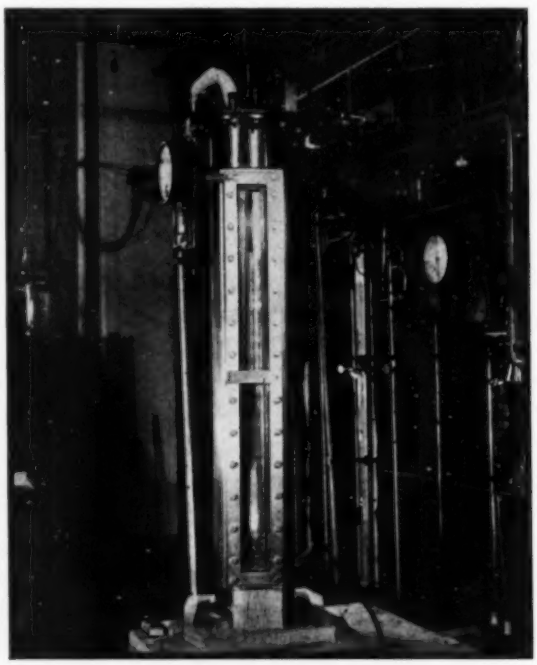


FIG. 11. The liquid helium apparatus at Ohio State (photograph by J. G. Daunt). Within the vertical window is the liquid helium container, including the bottle through whose ground-glass stopper the superfluid flow took place, enriching the residue with the light isotope of helium.

The last point to be mentioned in this section is not physics but instrumentation. It is the rapid development of solid counters for ionizing radiation. Everyone knows how many devices—ion chambers, proportional counters, Geiger counters—are available for the detection of ionizing radiation. All of these work on the same principle: the use of ionization in a gas. Ions so formed can be made to drift to a collecting electrode, sometimes very fast, and with secondary ionization to amplify the signal. But such counters have two disadvantages: they are large and rather inefficient for gamma-rays and very fast particles. Such radiation has such a large range in gas at normal pressure that not many ions are produced in a counter volume of convenient size. High pressure is troublesome, and the lowered mobility of the ions slows response. A fast counting device which was made of a solid, not a gas, would be just a thousand times smaller for a given atomic content, and might be fast.

Two such devices have been much worked on in 1947. One is really a solid ion chamber, the so-called crystal counter. A brilliant wartime thesis by a Dutch physicist, van Heerden, initiated this work.[32] It was an old but not very successful idea. He placed two electrodes across an insulating crystal. When ions were produced in the crystal, if the electrons had energy enough to go into one of the conduction bands of the crystal lattice, they could travel in the field between the electrodes and produce an amplifiable pulse. He used silver chloride, because it was known that it conducted no electric current, at least at liquid air temperatures, except when it was illuminated. The gamma- or alpha-ray ionizes the atom deep in the crystal just as ultraviolet does. But single pulses are counted, not the d.c. current. This device works well but it is still a little mysterious.

[32] P. J. van Heerden, *The Crystal Counter* (Dissertation, Utrecht, 1945).

Some diamonds, those which have a particular kind of ultraviolet absorption,[33] also appear to have the counting property. This ability to become temporarily conducting seems to depend upon some features of the crystal lattice which are easily disturbed by impurities and flaws. No doubt the electrons are set free with ample energy in every crystal, but only in a few can they travel to the electrodes. It seems that this development will indeed make for faster and more efficient counters when it has been a little better understood.[34]

The oldest method for detection of single fast particles is the spinthariscope. This is just a microscope used to watch a phosphorescent screen. When an alpha-particle hits the grain of phosphor, a large fraction of the kinetic energy appears as visible light. Rutherford did all his early fundamental work in a dark room, patiently counting the scintillations. A more modern adaptation of this idea is already providing a very useful counter. The microscope is omitted; the eye and hand are replaced by a photo-multiplier tube an amplifier, and scalar circuit; the zinc sulfide phosphor may be there, or it may be replaced by a rapid-responding fluorescent material, such as napthalene, plain mothballs.[35] The mothball counter is already proving a valuable tool, especially for counting gamma-rays with high efficiency. The solid counters are here to stay; they should become as familiar as the ubiquitous Geiger-Müller tube.

THE SUPERFLUID

The remarkable result of Ogg, who found that a solution of the alkali metals in liquid ammonia became a superconducting solid, was described at length in the review of 1946. While there has been one confirming report, several groups of experiments have failed to find Ogg's result. It must be said that this result is still uncertain. It is too bad that we do not yet have a superconductor which we can count on at such high temperatures as that of liquid air.

[33] Hofstadter, Milton, and Ridgeway, Phys. Rev. **72**, 977 (1947); L. F. Curtiss and B. W. Brown, Phys. Rev. **72**, 643 (1947); Friedman, Birks, and Gauvin, Phys. Rev. **73**, 186 (1948).

[34] F. Marshall and J. W. Coltman, Phys. Rev. **72**, 528 (A) (1947).

[35] M. Deutsch *et al.*, private communication.

FIG. 12. The soap-bubble model of a metal (photo from Sir L. Bragg). These are bubbles 0.76 mm in diameter formed from a very small glass jet. The picture shows the irregular "grain boundary" between three "microcrystals" of different crystal orientation. It is remarkably like a metal etch figure.

The year saw a new and interesting result in the long history of the superfluid liquid helium. It has long been known that below about 2°K liquid helium behaved in a unique matter. There is nothing else like it. For example, in the apparatus shown in Fig. 11 the material contained in the central flask simply leaked through an ordinary tight ground glass stopper. It is a liquid, but it has the viscosity of a gas. Many remarkable experiments have confirmed this general behavior.[36]

A good many years ago it was suggested that this behavior was to be ascribed to the strange properties of particles which have zero spin, and thus obey the Einstein-Bose statistics. It is known that a perfect Bose gas will form a condensed phase at low temperatures, entirely independent of any attractive forces between the particles. It was thought that this property was involved in the strange behavior of the superfluid.

A decisive test for this theory is the behavior of the light isotope of helium, He^3. This light helium is present in normal helium to about one part in ten million. Its atomic properties are identical

[36] Daunt, Probst, Johnston, Aldrich, and Nier, Phys. Rev. **72**, 502 (1947).

with those of ordinary helium, except in two respects. It is lighter, of course, and it has a non-zero spin, satisfying not the Einstein-Bose but the Fermi-Dirac statistics. Would this material become a superfluid?

The experiment was attempted this year, but not quite conclusively. It was indeed found that after the superfluid left the container, the remaining helium was enriched in the light isotope. The isotopic analysis, a very difficult problem for such small concentrations, was carried out by Nier at Minnesota, while the cryogeny was done by Daunt and at Ohio State.[36] The data seem unambiguous, but the conclusion is still uncertain. In the first place it is by no means clear that the material He^3 will behave normally in such high dilution; in the second place, the transition temperature for He^3 to go over to the superfluid form, if it in fact can, may lie lower than the temperatures so far employed. The really conclusive work can be done when macroscopic samples of light helium are available. Since it is formed by the decay of the radioactive isotope H^3, which can be made in quite large amounts by irradiation of lithium in the strong neutron flux of the chain reactors, we may hope that this experiment will be done soon. The answer to the problem of the superfluid will shed real light on the related questions of superconductivity, and on the properties of the other curious "quantum liquid," nuclear matter itself.

AN INTERESTING PAIR OF RESULTS

The main stream of physics research has a kind of style. The advances we have already described are the ones made in the fashionable fields, the ones thought of as fundamental, the newsiest ones. But it is worth while to spend a little time looking at two results which show the forever amazing freshness and the unexpected unity of the field of physics.

One such simple change in our ideas which has real consequences is a new measurement of the half-life of potassium 40, the radioactive isotope found in small concentration in ordinary potassium. The rather wide distribution of this isotope has always made it interesting, but its low activity, compared to uranium, has been regarded as removing from it much importance for the heat balance of the earth, and for general geological and biological questions. Measuring the activity of such a long-lived substance is the only means of measuring its half-life; we cannot observe decay in our lifetimes. Thus a beta-ray count from a weighed sample has been the source of our knowledge about the half-life of this substance. It was pointed out by two Norwegian radio-geologists[37] that the recent work of the Swiss physicists Bleuler and Gabriel[38] has important consequences for the history of the earth. The Swiss workers confirmed an earlier suspiscion: K^{40} decays in two ways. Some of its nuclei disintegrate by emitting a beta-particle, to become nuclei of calcium. These decays are easily identified by the rather fast beta-ray, and can be counted. On the basis of such decays the lifetime has been set at about 7×10^8 years. But the nucleus can decay also by capturing an atomic electron, to become an argon nucleus. This process sometimes results in gamma-rays, but most often in no detectable nuclear radiation at all. These decays then are missed, and falsify the measured lifetime. The work of Bleuler and Gabriel was to measure the x-rays which are emitted by the atom in filling up the "hole" after the K electron is swallowed by the nucleus. This gives a true lifetime of about 2.4×10^8 years. This sounds like a good piece of clean-up work, but hardly something of interest except for completing the table of radioactive isotopes. Not so. Gleditsch and Gráf remind us that potassium is responsible right now for only about a fifth of the heat produced in the earth's crust. But about 2.4 billion years ago, a time when the earth was new, the potassium must have produced a thousand times as much heat as it does now, while the elements we think of as the radioactive ones went along at less than double their present rate of heating. The heat caused by radioactivity in the earth's crust was hundreds of times greater then than now, and all from potassium. It is clear that early geology will be much affected by such an energy source. It is curious to note that living organisms of the sort we know today could not have evolved a few billion years ago; the potassium radioactivity was too great! The argon of the atmosphere originates from potassium decay.

---

[37] E. Gleditsch and T. Gráf, Phys. Rev. **72**, 640 (1947).
[38] E. Bleuler and M. Gabriel, Helv. Phys. Acta **20**, 67 (1947).

The last of the figures, Fig. 12, shows some soap bubbles. They have a story, too. At the Cavendish Laboratory Sir Lawrence Bragg and J. F. Nye have prepared a very ingenious and beautiful demonstration, from which real information can come.[39] The bubbles shown in the figure are small soap bubbles about $\frac{3}{4}$ of a millimeter in diameter, blown in the hundreds of thousands by a fine glass jet, and allowed to float as a "raft" of bubbles on the surface of the soapy water. Soap bubbles attract each other, as is well known, when they come close enough so that the capillary forces can work. But they do not coalesce; repulsive forces set in when the bubbles come so close that compression of the air content begins. The forces which join bubbles are not directional. Now, all these statements can be made about the forces between the atoms of a metal. For the first time a model has been prepared which has enough "particles" to give a fair idea of how a metal must act. And the result is spectacular. The figure here reproduced shows a grain boundary, with the microcrystals oriented differently in the two grains. If the reader will allow his imagination just a little scope, he can easily fancy himself looking at an etched metal. Notice the irregular and low density character of the grain boundary. Sometimes out-size bubbles, "impurity atoms," can be found concentrated in such boundaries. Slip, shear, impurity effects, elastic deformation, the "Beilby layer," all these and more phenomena have their analog in the bubble crystal. A three-dimensional model can be made by piling the bubbles in a thick layer; its properties are evidently more difficult to observe in detail. The many photos of the model which appear in the original paper are well worth looking up; it is a pity that the moving picture reels Sir Lawrence has taken are not widely available.

It seems appropriate to end with this simple idea. Physics includes very much; the line shift and the bubble model are two of the many and diverse events of physics in 1947. It is apparent to this reviewer that justice has been done to none of the events here described, and that many others were necessarily omitted. If a few readers will go to the original papers, or will recall for themselves valuable work not here reported, this review will have had some effect.

I am grateful for the kindness of those who supplied the figures reproduced here. Their names will be found in the figure captions.

[39] L. Bragg and J. Nye, Proc. Roy. Soc. **A190**, 474 (1947).

Calendar of Meetings

April

22–24 Acoustical Society of America, Washington, D. C.
22–24 American Philosophical Society, Philadelphia, Pennsylvania
25–30 American Ceramic Society, Incorporated, Chicago, Illinois
26–28 National Academy of Sciences, Washington, D. C.
26–28 Institute of Radio Engineers and the Radio Manufacturers Association (Spring Meeting), Syracuse, New York
28–30 American Institute of Electrical Engineers (Northeastern District Meeting), New Haven, Connecticut
29–May 1 American Physical Society, Washington, D. C.
30–May 1 Canadian Institute of Radio Engineers, Toronto, Ontario, Canada

May

3–5 International Scientific Radio Union and the Institute of Radio Engineers, Washington, D. C.
9–12 American Institute of Chemical Engineers, Cleveland, Ohio
17–21 Society of Motion Picture Engineers, Santa Monica, California
22 Institute of Radio Engineers (North Atlantic Region), Cambridge, Massachusetts
22 Society of Applied Spectroscopy, Brooklyn, New York
26–29 Canadian Association of Physicists (Third Annual Council), Ottawa, Canada
30–June 2 American Society of Refrigerating Engineers, Swampscott, Massachusetts
30–June 4 American Society of Mechanical Engineers (Semi-annual Meeting), Milwaukee, Wisconsin

June

6–11 Society of Automotive Engineers (Summer Meeting), French Lick, Indiana
14–18 American Society for Engineering Education, Austin, Texas
14–19 Symposium on Molecular Structure and Spectroscopy, Columbus, Ohio
21–23 American Physical Society, Madison, Wisconsin
21–25 American Institute of Electrical Engineers, Mexico City, Mexico
21–25 American Society for Testing Materials (Annual Meeting), Detroit, Michigan
24–26 American Physical Society, Pasadena, California
28–July 1 American Astronomical Society, Pasadena, California

Contributed Original Research

Self-Excited Mechanical Oscillations

N. MINORSKY
Stanford University, Stanford University, California
(Received July 11, 1947)

This paper deals with a phenomenon of self-excitation observed in an electromechanical system. It is shown that this phenomenon is expressible in terms of a difference-differential equation, having a frequency spectrum that depends on a parameter. The linear theory fails to give an answer to the question of stability of oscillations. It is necessary for this purpose to make use of the equations of the first approximation of the non-linear theory. Expressions for a stable stationary amplitude and frequency obtained on that basis show an agreement with the observed features of these self-excited oscillations.

1. FORMULATION OF THE PROBLEMS

THIS paper deals with certain self-excited oscillations observed in a mechanical system possessing a retarded action. The problem can be described as follows. Consider a system with one degree of freedom (e.g., a pendulum) acted on by an external periodic moment. Its differential equation is

$$J\ddot{\theta}+B\dot{\theta}+C\theta=M\sin\kappa t, \tag{1}$$

where J, B, and C are well-known physical constants, θ is the angle of oscillation, assumed to be small, and $M\sin\kappa t$ is the external moment. Very frequently, in order to increase the natural damping of the system, one provides an artificially produced moment, $S\dot{\theta}$, so that (1) is to be replaced by

$$J\ddot{\theta}+(B+S)\dot{\theta}+C\theta=M\sin\kappa t. \tag{2}$$

This is, for example, the underlying idea of the antirolling stabilization of ships where the moment $S\dot{\theta}$ is produced by properly timed displacements of water ballast between the tanks. As this problem has been treated elsewhere[1] we shall not enter into this matter here. In practice the coefficient S is a monotone increasing function $S(\lambda)$ of a certain parameter λ, the amplification factor of the electronic control in the problem just mentioned. It is generally observed that for a certain interval of λ, (2) describes the behavior of the system sufficiently accurately. If, however, λ exceeds a certain critical value, there appears an additional oscillation of parasitic nature which does not follow from (2). A cinematographic investigation has shown that this oscillation is apparently related to the existence of a time lag with which the moment $S\dot{\theta}$ is applied to the system. Since similar phenomena have been observed in other applications,[2] it was thought that it might be of interest to investigate this matter in more detail. Oscillations of this kind are also observed in the absence of any external excitation. For that reason it is convenient to consider instead of (2) the following difference-differential equation:

$$J\ddot{\theta}+B\dot{\theta}+S(\lambda)\dot{\theta}(t-h)+C\theta=0, \tag{3}$$

where $\dot{\theta}(t-h)=\dot{\theta}_h$ is the *retarded velocity*, h being the time lag which we shall assume to be constant. Dividing (3) by J and setting $B/J=p$; $S(\lambda)/J=q$ and $C/J=\omega_0^2$ we obtain

$$\ddot{\theta}+p\dot{\theta}+q(\lambda)\dot{\theta}_h+\omega_0^2\theta=0. \tag{4}$$

It is possible to ascertain the existence of an oscillatory process in which we are interested here on the basis of this equation, but as regards other features of these oscillations, such as their stability, stationary amplitudes, and so on, this equation is inadequate. For this purpose it is necessary to supplement the *linear* difference-

[1] N. Minorsky, Trans. A.S.M.E. **69**, 735 (1947).

[2] F. Rheinhardt, Wiss. Veröff. Siemens Werke (1936).

differential Eq. (4) by certain non-linear terms, and then the theory seems to be in agreement with the observed facts.

2. LINEAR PROBLEMS

We shall try to satisfy (4) by a solution of the form

$$\theta = \theta_0 e^{zt} = \theta_0 e^{(\alpha + j\omega)t}, \tag{5}$$

where α is either a decrement (if $\alpha<0$) or an increment (if $\alpha>0$) and ω is the frequency, where $j=\sqrt{-1}$. On substituting (5) into (4), multiplying by $(\alpha-j\omega)$, and setting $\omega h=\phi$, we have

$$\alpha+\frac{\alpha\omega_0^2}{\alpha^2+\omega^2}+p+qe^{-\alpha h}\cos\phi=0, \tag{6}$$

$$-\omega+\frac{\omega\omega_0^2}{\alpha^2+\omega^2}+qe^{-\alpha h}\sin\phi=0. \tag{7}$$

For a harmonic oscillation ($\alpha=0$), we have

$$p+q\cos\phi=0, \tag{8}$$

$$\frac{\omega_0^2-\omega^2}{\omega}+q\sin\phi=0. \tag{9}$$

Combining these equations, we obtain

$$(\omega_0^2-\omega^2)/p\omega=\tan\phi. \tag{10}$$

A harmonic oscillation can exist only if (8) and (10) are fulfilled simultaneously. We shall show that this is possible only for certain values of the parameter λ. In fact, (10) can be written as

$$(\omega_0^2h/p)(1/\omega h)-(1/ph)\omega h=\tan\phi.$$

Setting $\omega_0^2h/p=g$; $ph=n$, where g and n are positive constants, the preceding equation becomes

$$g/\phi-\phi/n=\tan\phi. \tag{11}$$

This transcendental equation can be solved graphically by finding the abscissas of the points of intersection of the two curves

$$y_1=g/\phi-\phi/n \quad \text{and} \quad y_2=\tan\phi.$$

The curve $y_1=y_1'+y_1''$ is obtained by adding the ordinates of a hyperbola $y_1'\phi=g$ and of a straight line $y_1''=-(1/n)\phi$ passing through the origin and having a negative slope $-1/n$. The graphical construction is shown in Fig. 1. The roots ϕ_0. ϕ_1, ϕ_2, $\cdots$ of (11) are at the points A, B, C, D, $\cdots$. The fact that (8) must also be satisfied imposes an additional restriction on these roots. In fact, since $p>0$, $q>0$ and $\cos\phi<0$, the regions in which $\cos\phi>0$ are ruled out so that only ϕ_1, ϕ_3, $\cdots$ in the regions where $\cos\phi<0$ are admissible. Since both (8) and (10) have to be satisfied simultaneously, it is apparent that this can occur only for certain discrete values of the parameter q for which the values of ϕ from (8) are precisely the values ϕ_1, ϕ_3, $\cdots$ obtained from (10) as shown in Fig. 1. We can call these particular values $q_1, q_3, \cdots$ of the parameter q, the *harmonic values* inasmuch as only when q reaches one of these values, say $q=q_1$, a harmonic oscillation ($\alpha=0$) with frequency $\omega_1=\phi_1/h$ is possible. If q differs from a harmonic value, the oscillation ceases to be harmonic ($\alpha\neq0$). In what follows we will attach subscripts to indicate the harmonic values, e.g., q_1, ω_1, ϕ_1; q_3, ω_3, ϕ_3; $\cdots$, corresponding to the different modes of harmonic oscillations. The most important of these modes is generally the first one, and we will confine our attention only to this first mode. Although it is possible to obtain higher modes, their amplitudes are generally so small as to be of no interest. As regards the first mode, it is clear that ϕ_1 may be either in the second quadrant ($\pi/2<\phi_1<\pi$) or in the third ($\pi<\phi_1<3\pi/2$), depending upon the shape of the curve y_1, that is, ultimately, on the constant parameters p, h, and ω_0^2 of the system. If ϕ_1 is in the second quadrant, $\sin\phi_1>0$ and from (9) it follows that $\omega_1^2>\omega_0^2$; if ϕ_1 is in the third quadrant, $\omega_1^2<\omega_0^2$.

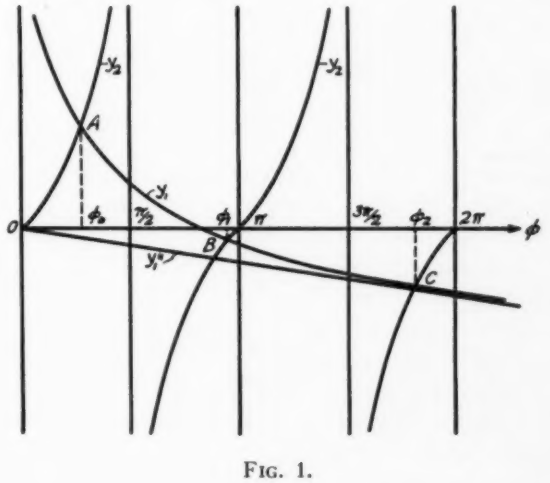


FIG. 1.

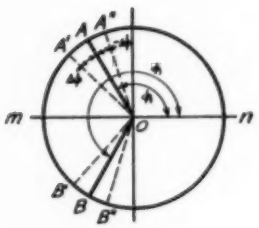


FIG. 2.

For reasons which will appear later we shall be interested only in solutions (5) not far from the harmonic solutions ($\alpha=0$). In that neighborhood α is small. The substitution of (5) into (4) gives:

$$z+(\omega_0^2/z)+p+qe^{-hz}=0. \tag{12}$$

Let us consider a small change Δq of the parameter from its harmonic value q_1 for which $z=j\omega_1$. For $q_1+\Delta q$, the corresponding value of z will be

$$z=j\omega_1+\Delta z=j\omega_1+\Delta\alpha+j\Delta\omega.$$

Substituting this value in (12) and carrying out calculations to the first order of small quantities Δq, $\Delta\alpha$, and $\Delta\omega$, one gets

$$\Delta z=\Delta\alpha+j\Delta\omega=-\Delta q/(m^2e^{j\phi_1}-hq_1)\,;\ m^2=1+\omega_0^2/\omega_1^2$$

The real and the imaginary parts of this expression yield

$$\Delta\alpha/\Delta q=-(m^2\cos\phi_1-hq_1)/(m^4-2q_1hm^2\cos\phi_1+h^2q_1^2),$$

$$\Delta\omega/\Delta q=(m^2\sin\phi_1)/(m^4-2q_1hm^2\cos\phi_1+h^2q_1^2).$$

Since $\cos\phi_1=-p/q$; $\sin\phi_1=\pm(q_1^2-p^2)^{\frac{1}{2}}/q_1$ these expressions become

$$\Delta\alpha/\Delta q=(m^2p+hq_1^2)/[q_1(m^4+2phm^2+h^2q_1^2)], \tag{13}$$

$$\Delta\omega/\Delta q=\pm(m^2(q_1^2-p^2)^{\frac{1}{2}})/[q_1(m^4+2phm^2+h^2q_1^2)]. \tag{14}$$

Since the right-hand side of (13) is always positive, $\Delta\alpha>0$ for $\Delta q>0$ and *vice versa* but, since for $\Delta q=0$, $\alpha=0$, this means that for a small increase ($\Delta q>0$) of q from its harmonic value, $\alpha>0$ and for a decrease ($\Delta q<0$), $\alpha<0$. The sign of the right-hand side of (14) depends on that of $\sin\phi_1$; if ϕ_1 is in the second quadrant, the plus sign must be used; if it is in the third quadrant, one must take the minus sign. Since $\Delta\phi=h\Delta\omega$, the same remark applies to $\Delta\phi/\Delta q$. These conclusions can be summarized in connection with Fig. 2 for the two cases just mentioned, *viz.*: $\pi/2<\phi_1<\pi$ (point A) and $\pi<\phi_1<3\pi/2$ (point B). In the first case for $\Delta q>0$, $\Delta\phi>0$ (point A') and for $\Delta q<0$, $\Delta\phi<0$ (point A''). In the second, for $\Delta q>0$, $\Delta\phi<0$ (point B') and for $\Delta q<0$, $\Delta\phi>0$ (point B''). Hence, in both cases for $\Delta q>0$, the phase angle shifts into the sector AmB (points A', B') and for $\Delta q<0$, the phase angle (A'', B''), on the contrary, is shifted into the sector AnB. We may call the sector AmB, the *sector of self-excitation*, inasmuch as for phase angles ϕ in this sector $\alpha>0$ and the motion occurs with increments. In the sector AnB, $\alpha<0$ and the oscillation dies out.

The harmonic thresholds themselves (points A and B) appear thus as *divides* separating the values of ϕ for which $\alpha>0$ from those for which $\alpha<0$. Using the terminology of Poincaré,[3] the harmonic value q_1 appears as the *bifurcation value* for the parameter q. For $q=q_1-\Delta q$, the differential equation has a stable focal point; for $q=q_1+\Delta q$ the focal point is unstable and for $q=q_1$, the threshold itself, the singularity is a vortex point and the system behaves as a harmonic oscillator. On the basis of the linear theory the amplitudes increase indefinitely if $\alpha>0$, which is contrary to the observed facts. It is necessary, therefore, to adopt a non-linear treatment of the problem.

3. NON-LINEAR PROBLEM

Instead of (4) we will consider now a quasi-linear difference-differential equation

$$\ddot{\theta}+p\dot{\theta}+q(\lambda)\dot{\theta}_h+\omega_0^2\theta+\mu f(\dot{\theta},\dot{\theta}_h)=0, \tag{15}$$

where $f(\dot{\theta},\dot{\theta}_h)$ is a non-linear function of $\dot{\theta}$ and $\dot{\theta}_h$ and μ is a small parameter. The condition of the quasi-linearity restricts the investigation to neighborhoods of the periodic solutions ($\alpha=0$) of the linear problem ($\mu=0$). This is also a case which is of interest in practice since the non-linearities are generally small in a well designed installation. If one succeeds in showing that there exists a closed solution of (15) of the "limit cycle" type

[3] H. Poincaré, *Figures d'Equilibre d'une Masse Fluide* (Paris, 1903); A. Andronow and Chaikin, *Teorie of Oscillations* (Moscow, 1937); Translation of this book by S. Lefschetz will appear shortly in the Princeton University Press.

in the neighborhood of the harmonic solution, the problem is solved. In general it is not certain that a solution of this kind exists. It may exist, however, for some special forms of the non-linear function f, and this will be the subject of our inquiry.

Following the Kryloff-Bogoliuboff method[4] we shall look for a solution of the form

$$\theta = a\sin(\omega t+\beta), \tag{16}$$

where a, the amplitude, and β, the phase angle, are certain unknown functions of time. Moreover, it is desired that $\dot\theta$ should be of the form

$$\dot\theta = a\omega\cos(\omega t+\beta), \tag{17}$$

which imposes an additional condition

$$\dot a\sin(\omega t+\beta)+a\dot\beta\cos(\omega t+\beta)=0. \tag{18}$$

Differentiating $\dot\theta$, considering a and β as variables, substituting θ, $\dot\theta$, and $\ddot\theta$ in (15), and setting $\omega t+\beta=\gamma$ one obtains

$$\dot a+a[(\omega_0^2-\omega^2)/\omega+q\sin\phi]\sin\gamma\cos\gamma +a(p+q\cos\phi)\cos^2\gamma+(\mu/\omega)f\cos\gamma=0, \tag{19}$$

$$\dot\beta a-a(p+q\cos\phi)\sin\gamma\cos\gamma -a[(\omega_0^2-\omega^2)/\omega+q\sin\phi]\sin^2\gamma -(\mu/\omega)f\sin\gamma=0, \tag{20}$$

where $f=f(a\omega\cos\gamma, a\omega\cos(\gamma-\phi))$. In this expression for f it is assumed that the retarded amplitude $\bar a$ is equal to the non-retarded one a, which seems to be plausible in the first approximation if the time lag is small. These equations are of the quasi-linear type if the problem is restricted to a small neighborhood around ω_1 and ϕ_1, in which case the quantities $(\omega_0^2-\omega^2)/\omega+q\sin\phi=\epsilon$ and $p+q\cos\phi=\eta$ are small, which we will assume. This means that during one period $T=2\pi/\omega$ of the trigonometric functions the quantities a and β vary very little. One can apply, therefore, the standard procedure[5] of "averaging" per period T which leads to the following equations

$$\frac{a(t+T)-a(t)}{T}+\frac{a\epsilon}{2\pi}\int_0^{2\pi}\sin\gamma\cos\gamma d\gamma +\frac{a\eta}{2\pi}\int_0^{2\pi}\cos^2\gamma d\gamma+\frac{\mu}{2\pi\omega} \times\int_0^{2\pi}f(a\omega\cos\gamma, a\omega\cos(\gamma-\phi))\cos\gamma d\gamma=0,$$

$$\frac{\beta(t+T)-\beta(t)}{T}-\frac{\eta}{2\pi}\int_0^{2\pi}\sin\gamma\cos\gamma d\gamma -\frac{\epsilon}{2\pi}\int_0^{2\pi}\sin^2\gamma d\gamma-\frac{\mu}{2\pi a\omega} \times\int_0^{2\pi}f(a\omega\cos\gamma, a\omega\cos(\gamma-\phi))\sin\gamma d\gamma=0.$$

If one considers the slow variations of the amplitude and phase in the course of many periods T, one can replace $a(t-T)-a(t)$ by Δa, etc., and assume T of the previous problem as ΔT of the present one. In the first approximation one can write, therefore, the first terms in the preceding equations as da/dt and $d\beta/dt$, but it must be borne in mind that these new notations relate now to the averaged process occurring in the course of many periods of the trigonometric functions.[5] With these notations the equations of the first approximation become

$$\frac{da}{dt}=-\frac{a\eta}{2}-\frac{\mu}{2\pi\omega} \times\int_0^{2\pi}f(a\omega\cos\gamma, a\omega\cos(\gamma-\phi))\cos\gamma d\gamma, \tag{21}$$

$$\frac{d\beta}{dt}=\frac{\epsilon}{2}+\frac{\mu}{2\pi a\omega} \times\int_0^{2\pi}f(a\omega\cos\gamma, a\omega\cos(\gamma-\phi))\sin\gamma d\gamma. \tag{22}$$

These equations differ from the Kryloff-Bogoliuboff equations[6] only by the terms with ϵ and η arising from the transcendental character of this problem.

[4] N. Kryloff and N. Bogoliuboff, *Introduction to Non-Linear Mechanics* (Kieff, 1937); English translation by S. Lefschetz, Princeton University Press, Princeton, New Jersey, 1943); also N. Minorsky, *Introduction to Non-Linear Mechanics* (Edwards Brothers, Ann Arbor, Michigan, 1947), p. 187.

[5] Reference 4; also B. van der Pol, Phil. Mag. **7**, 3, 35 (1927); Phil. Mag. **6**, 43 (1922).

[6] Kryloff-Bogoliuboff, reference 4; N. Minorsky, reference 4, p. 188.

It is necessary now to restrict the problem to the case with which we were concerned in these experiments. The artificially produced moment $S(\lambda)\dot{\theta}_h$ was obtained in our experiments by displacing a certain amount of water ballast by means of the axial variable pitch pumps controlled through an electronic system from instruments responsive to the angular motion of the system. Not entering into the technical details of this arrangement described elsewhere,[1] it is sufficient to mention here that, owing to the inherent limitation of pumps and driving equipment, the transferred ballast for larger $\dot{\theta}_h$ was slightly less than in accordance with the linear law, thus exhibiting a somewhat "drooping" characteristic with the load. One can, therefore, assume that instead of $q\dot{\theta}_h$, the real characteristic is here of the form $q\dot{\theta}_h - q_3\dot{\theta}_h^3$, where the coefficient q_3 is very small as compared to q which, incidentally, enables us to dispense with the use of the small parameter μ explicitly. Since the term $q\dot{\theta}_h$ has already been taken into account in the preceding equations, we have

$$f(\dot{\theta}, \dot{\theta}_h) = f(\dot{\theta}_h) = -q_3\dot{\theta}_h^3. \tag{23}$$

Taking

$$\dot{\theta}_h = a\omega\cos(\omega t+\beta-\phi) = a\omega\cos(\gamma-\phi)$$

and replacing it in (21) with the form (23) for f, one obtains, after the averaging the following expression,

$$\frac{da}{dt} = -\frac{a(p+q\cos\phi)}{2} + \frac{3}{8\omega}q_3a^3\omega^3\cos\phi. \tag{24}$$

The stationary amplitude a_{sr} is thus given by the expression

$$a_{sr} = (1/\omega_{sr})(4(p+q\cos\phi_{sr})/3q_3\cos\phi_{sr})^{\frac{1}{2}}. \tag{25}$$

Since $\cos\phi$ is always negative, the stationary amplitude can exist only when the quantity $p+q\cos\phi_{sr}<0$, that is, when the phase angle ϕ_{sr} is in the sector of self-excitation.

It is interesting to note that for $q=q_1$ and for the linear case ($q_3=0$) the stationary amplitude given by (25) is indeterminate since $p+q_1\cos\phi_1=0$. In fact, in such a case the system becomes an ordinary harmonic oscillator capable of oscillating with any amplitude, depending on the initial conditions. In the non-linear case, on the contrary, the stationary amplitude is determined uniquely by the parameters of the system and does not depend on the initial conditions.[7]

4. STATIONARY AMPLITUDE AND PHASE ANGLE

The problem reduces to the calculation of the stationary amplitude a_{sr} from (25), but since this expression depends on the stationary phase angle ϕ_{sr}, it is necessary to determine first the latter. Setting $\phi_{sr}=\phi_1+\Delta\phi_{sr}$, where $\Delta\phi_{sr}$ is the stationary phase-angle correction, it is clear that

$$\Delta\phi_{sr} = \Delta\phi + \delta\phi, \tag{27}$$

where $\Delta\phi = h\Delta\omega$ is given by (14), *viz.*:

$$\Delta\phi = \pm hm^2(q_1^2-p_1^2)^{\frac{1}{2}}\Delta q/ q_1(m^4+2hpm^2+h^2q_1^2), \tag{28}$$

with $\Delta q>0$. This component correction $\Delta\phi$ is, therefore, the phase shift in the linear problem, as was explained in Section 2. The other component, $\delta\phi$, is the non-linear phase correction which can be calculated from (22). Carrying out the averaging with f given by (23), and introducing a new variable $\psi=\omega t+\beta$, one obtains the following expression

$$\frac{d\psi}{dt} = \Omega(a) = \omega + \tfrac{1}{2}\left[\frac{\omega_0^2-\omega^2}{\omega} + q\sin\phi - \tfrac{3}{4}q_3\sin\phi\cdot a^2\omega^2\right] = \omega+\delta\omega.$$

In this expression $\Omega(a)$ is the non-linear frequency which becomes stationary if $a=a_{sr}$.

Replacing in this equation ω by $\omega_1+\Delta\omega$, one obtains

$$\delta\phi = \frac{h}{2}\left[\frac{\omega_0^2-(\omega_1+\Delta\omega)^2}{\omega_1+\Delta\omega} + (q_1+\Delta q)\sin(\phi_1+\Delta\phi) - \tfrac{3}{4}q_3a_{sr}^2(\omega_1+\Delta\omega)^2\sin(\phi_1+\Delta\phi)\right].$$

The quantities Δq, $\Delta\omega$, and $\Delta\phi=h\Delta\omega$ will be considered as small and of the first order, and the calculation is carried out only to that order. This gives

$$2\delta\phi = \Delta\phi\left[-\frac{(\omega_0^2+\omega_1^2)}{\omega_1^2} + hq_1\sin\phi_1\right]. \tag{29}$$

[7] Andronow and Chaikin, reference 3, Chapter V; also N. Minorsky, reference 4, Chapter IV.

The same procedure applied to (25) gives

$$a_{sr}^2 \cong (4/3\omega_1^2)(\Delta q/q_3 - \Delta\phi_{sr}(q_1/q_0)\tan\phi_1),$$

that is,

$$(3/4)q_3\omega_1^2 a_{sr}^2 = \Delta q(1-(\Delta\phi_{sr}/\Delta q)q_1\tan\phi_1), \quad (30)$$

where

$$(\Delta\phi_{sr}/\Delta q) = (\Delta\phi+\delta\phi)/\Delta q$$
$$= (\Delta\phi/\Delta q)+(\delta\phi/\Delta\phi_{sr})(\Delta\phi_{sr}/\Delta q),$$

so that

$$(\Delta\phi_{sr}/\Delta q) = (\Delta\phi/\Delta q)[1-(\delta\phi/\Delta\phi_{sr})].$$

Replacing the last expression in (30) and $\Delta\phi/\Delta q$ from (14) one obtains the following expression

$$a_{sr} = (1/\omega_1)[(4\Delta q/3q_3) \times (1-(\Delta\phi_{sr}/\Delta q)q_1\tan\phi_1]^{\frac{1}{2}}, \quad (31)$$

where

$$\frac{\Delta\phi_{sr}}{\Delta q} = \frac{2hpm^2\sin\phi_1}{(m^4+2phm^2+h^2q_1^2)(2p+m^2p+hq_1^2)}. \quad (32)$$

If one substitutes (32) into (31), one notes that $\sin\phi_1\tan\phi_1<0$ for any ϕ_1 in the sector of self-excitation so that the parenthesis in (31) can be written as $(1+\gamma^2)$ and (31) becomes

$$a_{sr} = (1/\omega_1)[(4\Delta q/3q_3)(1+\gamma^2)]^{\frac{1}{2}}, \quad (33)$$

where $\gamma^2 = -[(\Delta\phi_{sr}/\Delta q)q_1\tan\phi_1]$. Equation (33) gives the stationary amplitude in the first approximation as a function of harmonic values (ω_1, ϕ_1, q_1) and constant parameters (ω_0, h, p). It is seen that a_{sr} exists only if $\Delta q>0$. On the other hand, as was shown by Eq. (13), the same condition $\Delta q>0$ results in the appearance of an unstable singularity. Thus we find the familiar features of the bifurcation theory of Poincaré[8] *viz.:* for $\Delta q<0$ the singularity is stable and there is no limit cycle; for $\Delta q>0$ the singularity is unstable and the limit cycle exists; the threshold $q=q_1$ (i.e., $\Delta q=0$) is the bifurcation value of the parameter q.

5. SELF-EXCITATION AND STABILITY

If the self-excitation develops from rest, initially $a=a_0\cong 0$, so that the second term on the right side of (24) can be neglected and the integration yields

$$a = a_0\exp[-\tfrac{1}{2}(p+q\cos\phi)t].$$

The oscillation can start from rest only if $p+q\cos\phi<0$. Replacing in this expression q by $q_1+\Delta q$, ϕ by $\phi_1+\Delta\phi$, and proceeding as before, one obtains

$$da/dt = (a_0/2)[\pm(q_1^2-p^2)^{\frac{1}{2}}\Delta\phi+(p/q_1)\Delta q].$$

If $\sin\phi_1>0$ one has to take the plus sign before the square root, and if $\sin\phi_1<0$, the minus sign must be taken. In the first case for $\Delta q>0$, $\Delta\phi>0$, and in the second for $\Delta q>0$, $\Delta\phi<0$, so that in both cases $(da/dt)_{a=a_0}>0$. If, however, $\Delta q<0$, in both cases $(da/dt)_{a=a_0}<0$ so that there is no self-excitation.

Finally, in order to ascertain whether the stationary amplitude is stable, it is necessary to differentiate the right-hand side of (24), *viz.:*

$$\psi(a) = -(a/2)(p+q\cos\phi)+(3/8\omega)q_3a^3\omega^3\cos\phi,$$

with respect to a and to substitute $a=a_{sr}$ after the differentiation. If $\psi_a(a_{sr})<0$ the stationary amplitude is stable; if $\psi_a(a_{sr})>0$, it is unstable. In this case one has

$$\psi_a(a_{sr}) = p+q\cos\phi_{sr}<0.$$

The limit cycle is, therefore, stable. This is also in agreement with observation which shows that these oscillations develop from rest and are stable.

6. CONCLUDING REMARKS

It may be of interest to say a few words concerning the physical nature of these phenomena. It was shown that the "harmonic threshold" separating the zone of self-excitation from that in which no self-excitation exists is given by (8) or, graphically, by *CA* and *OB* in Fig. 2. Since the constant p is a measure of the dissipative moment and $q\cos\phi$ is the in-phase component of the external moment, condition (8) means that the dissipative moment is just balanced by this in-phase component of the external moment so that the system behaves as a harmonic oscillator.

A condition of this kind is well known in connection with electron tube circuits, for example, the case when the feedback in a regenera-

[8] H. Poincaré, reference 3; also N. Minorsky, reference 4, Chapter V.

tive amplifier is increased to a critical value at which the system is on the threshold between the amplifier and the oscillator actions. At this threshold the system is not yet self-excited, but if the oscillation is started by some impulse it will, at least theoretically, persist indefinitely with a constant amplitude.

If the value of the parameter is increased above the harmonic value $q=q_1$, more energy is communicated to the system than it can dissipate initially. The amplitudes will grow until the average balance per cycle between the energies absorbed and dissipated is ultimately established. The amplitude at which this balance occurs will then become the *stationary amplitude*. This ultimate balance is possible only in the case when the system is non-linear; for a linear system no such balance exists for any amplitude, however large. For this reason any self-excited system which ultimately approaches a stationary state is necessarily non-linear. If the non-linearity is very small and the system is self-excited, the stationary amplitude is large, as is seen from (24). Physically this means that the amplitude will increase up to a sufficiently large value for which the average balance per cycle between the energies absorbed and dissipated is reached. In view of this it would be erroneous to attempt to reduce the non-linearity of a self-excited system with a view to eliminating these parasitic oscillations, since this merely increases the value of the stationary amplitude, as is apparent from (25). A correct procedure is to prevent the occurrence of self-excitation in general.

These considerations apply to the mechanical self-excited oscillations studied here as well as to similar oscillations in electron tube circuits. There exists, however, an essential difference between these two classes of non-linear oscillations. A great majority of electron tube circuits, at least those which do not exhibit quasi-discontinuities at some points of their cycle, are reducible to the differential equation of van der Pol. This means that for μ very small the self-excited system oscillates with a frequency not far from the natural frequency of the system. In our case for very small μ the self-excited oscillations stabilize around a frequency ω_1, which may be far from ω_0, but generally depends on ω_0 as well as on three additional parameters p, q, and h. This results from the fact that for $\mu=0$ the differential equation does not reduce to the equation of a simple harmonic oscillator with frequency ω_0, as is the case for equations of the van der Pol type, but reduces to a more complicated equation with a transcendental characteristic equation, as was outlined in Section 2.

The writer is indebted to Dr. E. Pinney, University of California, for valuable discussions of this matter.

This work was carried out under research programs of the Bureau of Ships (D. Taylor Model Basin) and the Office of Naval Research, Navy Department.

Erratum: A Magnetron Resonator System

[J. App. Phys. **18**, 1098 (1947)]

E. C. Okress
Electronics Engineering Department, Westinghouse Electric Corporation, Bloomfield, New Jersey

THE left-hand captions in Tables I and II should read "ratio of final to initial parameter" instead of the sequence indicated.

In relations (6a) and (6b) the unidentified coefficient

$$B_m=B\frac{d}{Y}\frac{k}{k_{xm}X}\frac{\sin\left(k_{y0}+\frac{2\pi m}{Y}\right)\frac{d}{2}}{\left(k_{y0}+\frac{2\pi m}{Y}\right)\frac{d}{2}}.$$

An Electronic Simultaneous Equation Solver

EDWIN A. GOLDBERG AND GEORGE W. BROWN*
Radio Corporation of America Laboratories, Princeton, New Jersey
(Received August 5, 1947)

An analogy device for directly solving simultaneous linear equations (without resorting to successive approximation methods) is described in this paper. It utilizes a number of high gain amplifiers interconnected by networks whose elements bear definite relationships to the known coefficients and constants of the system of equations. The stability criterion for such a system, and its application to an equation solver are described. Since the device produces the answers to a system of equations without delay, it is readily adaptable to problems of synthesis as well as those of analysis.

THE solutions of many problems which arise in the field of engineering necessitate the solving of systems of linear simultaneous equations. Systems involving up to three equations can normally be solved quite readily by straight elimination methods. These methods become quite cumbersome, however, when the number of equations exceeds three. An electronic device which will solve systems of equations up to ten in number has been constructed at RCA Laboratories, and will be described in this paper. The principles employed can be used to construct a device for solving systems of any number of simultaneous linear equations, and the same basic circuit design as employed in the ten equation solver can be used practically in devices for solving up to about twenty-five simultaneous equations.

The basic principle employed in the equation solver is quite simple.

Figure 1 is a block diagram of a device for solving a system of two equations. The system to be solved is as follows:

$$Y_1 = a_{1-1}X_1 + a_{1-2}X_2, \tag{1}$$

$$Y_2 = a_{2-1}X_1 + a_{2-2}X_2. \tag{2}$$

The values of the Y's and a's are known, and it is desired to determine the values of the X's. Assume that the output voltages of the two amplifiers are X_1 and X_2, respectively. These output voltages are divided by potentiometers set proportional to the a's, and their outputs fed back to the amplifier inputs through resistors of value R. The output of a signal generator (1000 cycles in the actual model) is applied to potentiometers P_1, P_2, and P_3. The ratios of potentiometers P_1 and P_2 are set proportional to the Y_1 and Y_2 values. P_3 is used in a bridging circuit to facilitate very accurate measurement of the answers. The value R is very large relative to the potentiometer resistances, so that loading effects will be negligible. ϵ_1 and ϵ_2 are the input voltages of amplifiers No. 1 and No. 2, respectively. $-G$ is the gain of the amplifiers. The equations for the system are as follows:

$$\epsilon_1 = \frac{a_{1-1}X_1 + a_{1-2}X_2 - Y_1}{3}, \tag{3}$$

$$\epsilon_2 = \frac{a_{2-1}X_1 + a_{2-2}X_2 - Y_2}{3}, \tag{4}$$

$$X_1 = -G\epsilon_1 \tag{5}$$

$$X_2 = -G\epsilon_2. \tag{6}$$

Substituting (5) and (6) into (3) and (4), and transposing, the equations for the system become:

$$Y_1 = \left(a_{1-1} + \frac{3}{G}\right)X_1 + a_{1-2}X_2, \tag{7}$$

$$Y_2 = a_{2-1}X_1 + \left(a_{2-2} + \frac{3}{G}\right)X_2. \tag{8}$$

If the factor $3/G$ is made negligible relative to the diagonal terms a_{1-1} and a_{2-2}, or if the value $3/G$ is subtracted from the diagonal terms, (7) and (8) become identical to (1) and (2). In practice, $3/G$ is made very small by using high gain amplifiers so that it may be neglected, and so that minor variations in amplifier gains will

* Now with the Statistical Laboratory, Iowa State College, Ames, Iowa.

not affect the accuracy of solution. The values of X_1 and X_2 are measured accurately by comparing their values against the signal generator voltage by means of balancing the measuring bridge with P_3 and reading the value of the setting of P_3. X_1 and X_2 could be measured somewhat less accurately by measuring the voltage outputs of the amplifiers with voltmeters. The level of the signal generator would then have to be accurately adjusted to insure accuracy.

Provisions for handling negative quantities were not included in Fig. 1. Negative quantities are taken care of by providing push-pull outputs for the signal generator and amplifiers, and making provisions to switch the potentiometers to either output side.

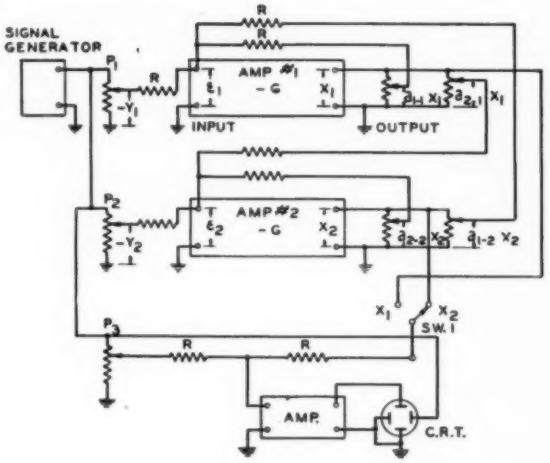


FIG. 1. Block diagram for a two equation computer.

STABILITY

The problem of stable operation arises whenever feedback is applied to an amplifier. Since the equation solver is a system of amplifiers with multipath feedback, the stability problem may, in some cases, become quite troublesome. This problem has been thoroughly investigated, and a brief description of the investigation and results follows:

For an n equation system, the equations for the response of the system are as follows:

$$Y_1=\left(a_{1-1}+\frac{n+1}{G(P)}\right)X_1+a_{1-2}X_2 +\cdots+a_{1-n}X_n, \quad (9)$$

$$\vdots$$

$$Y_n=a_{n-1}X_1+a_{n-2}X_2 +\cdots+\left(a_{n-n}+\frac{n+1}{G(P)}\right)X_n.$$

Let

$$\frac{n+1}{G(P)}=-\lambda, \quad (10)$$

where $G(P)$ is the gain characteristic[1] in operator form, and (9) becomes

$$\begin{aligned} Y_1&=(a_{1-1}-\lambda)X_1+a_{1-2}X_2+\cdots a_{1-n}X_n,\\ &\vdots\\ Y_n&=a_{n-1}X_1+a_{n-2}X_2+\cdots(a_{n-n}-\lambda)X_n. \end{aligned} \quad (11)$$

[1] The mathematical expression relating the gain (phase and magnitude) and angular frequency, ω. To obtain the operator form, replace every $j\omega$ by the differential operator P.

The determinant itself is the only factor in the system which need be investigated. It is,

$$\begin{vmatrix} (a_{1-1}-\lambda) & a_{1-2}\cdots a_{1-n} \\ \vdots & \\ a_{n-1} & a_{n-2}\cdots(a_{n-n}-\lambda) \end{vmatrix}. \quad (12)$$

The exponents of the terms in the indicial admittance (response to a unit step) will be the values of the P's in Eq. (10) if (12) is set equal to zero, solved for λ's, and the λ's substituted in (10), and (10) solved for P's. If any one of the P's has a positive real part, the system will be unstable, since a small impulse will cause a response which will increase in magnitude as time increases.

The amplifiers employed were so designed that their phase shift did not exceed $\pm 90°$ unless the gain were less than unity. With such a design, the system will be stable as long as all the λ's (known as the characteristic roots) have positive real parts or zero real parts. This condition is fulfilled by the equations for stable electrical networks if they are written down in a mesh order such that the resulting determinant is symmetrical about the diagonal, and a very large class of equations encountered in other physical problems.

The characteristic roots of a system are changed if rows or columns are interchanged, or if one or more rows or columns is multiplied by a constant. An unstable configuration is immediately evident upon examining the cathode-ray

tube bridge balance indicator, which will indicate circuit oscillation with no input.

Equation solver circuits which become unstable when a determinant is inserted may always be made stable if the variables are all finite. Using matrix notation, the system of equations,

$$\begin{aligned} Y_1 &= a_{1-1}X_1 + a_{1-2}X_2 + \cdots a_{1-n}X_n, \\ &\vdots \\ Y_n &= a_{n-1}X_1 + a_{n-2}X_2 + \cdots a_{n-n}X_n, \end{aligned} \tag{13}$$

may be represented by,

$$Y = A \cdot X. \tag{14}$$

The values of the characteristic roots of A specify the condition of the system with respect to stability. These may be made positive real by performing the following operation to obtain a new determinant,

$$A^{tr}A, \tag{15}$$

where A^{tr} is A with rows and columns interchanged. The system of equations then becomes

$$A^{tr}Y = A^{tr}AX, \tag{16}$$

and this system will always be stable. This transformation may be readily done on a desk calculator.

The above operation may be done without any calculation if the number of equations involved does not exceed half the number the equation solver can normally handle. The Y vector, A matrix, and the $-A^{tr}$ matrix are set in the equation solver as follows:

$$\begin{pmatrix} 0 \\ Y \end{pmatrix} \begin{pmatrix} I & -A^{tr} \\ A & 0 \end{pmatrix} \tag{17}$$

and the answers read as before. I is a unit matrix, and 0 is a zero matrix. Normally, the results are very insensitive to $-A^{tr}$, I, and the 0 vector above the Y vector, except in borderline cases. The operation above could be done in one operation by ganging appropriate controls in an equation solver which would be stable regardless of the manner in which the equations were set up on it.

OPERATION

Unity is the largest value for a coefficient (a) or constant term (Y) that may be set in the equation solver. A system of equations must be reduced accordingly before being set up in the computer.

A system of three equations, as follows, will be used to illustrate the process of reduction.

$$\begin{aligned} Y_1 &= a_{1-1}X_1 + a_{1-2}X_2 + a_{1-3}X_3, \\ Y_2 &= a_{2-1}X_1 + a_{2-2}X_2 + a_{2-3}X_3, \\ Y_3 &= a_{3-1}X_1 + a_{3-2}X_2 + a_{3-3}X_3. \end{aligned} \tag{18}$$

The determinant of the above system is

$$\begin{vmatrix} a_{1-1} & a_{1-2} & a_{1-3} \\ a_{2-1} & a_{2-2} & a_{2-3} \\ a_{3-1} & a_{3-2} & a_{3-3} \end{vmatrix}. \tag{19}$$

The constant terms are

$$\begin{matrix} Y_1 \\ Y_2 \\ Y_3. \end{matrix} \tag{20}$$

In order to realize the best direct results from the instrument, proceed according to the following outline:

Step number	Operation	Answer multiplier to make results check original equations
1	Multiply both sides of each equation by the same constant so that the largest coefficient (an a) in each equation becomes 1 (or slightly less).	Step 1 has no effect on the X's.
2	Multiply each column of coefficients (in the determinant) by a constant (K_1 for column 1, K_2 for column 2, etc.) so chosen that the largest coefficient in each column becomes unity (or slightly less). Note: This step is not essential, but is frequently worth while since it has a tendency to make the instrument operate at the optimum scale factors and thus produce most accurate direct results.	Multiply value obtained for X_1 by K_1, X_2 by K_2 etc., to satisfy original equations
3	Multiply the column of constants (Y's) by a factor (F) such that the largest Y becomes unity (or slightly less).	Divide all values obtained for X's by F to satisfy original equations.

Having reduced the system to a suitable form for the equation solver, coefficients of the X's are set on the dials. The X switch is set to X_1, and

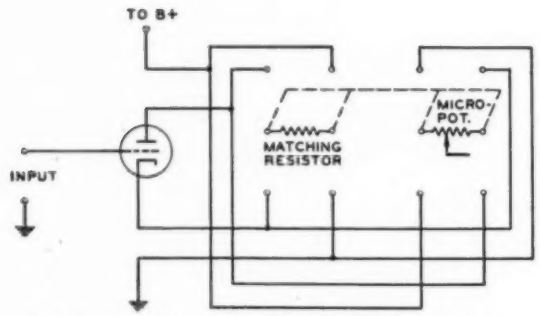


FIG. 2. Basic schematic of push-pull output stage.

the cathode-ray tube observed. If the pattern indicates stability (no oscillation) then the constant terms (Y's) may be set on their dials, signal introduced, and the answers read by setting the selector switch to the desired unknown, balancing the bridge for minimum vertical deflection, and reading the answer. A precaution to observe is not to introduce enough signal to cause overloading of any amplifier. Overloading can generally be detected from the appearance of distortion in the balanced pattern. The signal level at which the balances (value of the unknowns) change with level is the level where overloading occurs.

If it is not possible to obtain a balanced condition for an unknown, its value is larger than ten. The procedure to follow in such a case is to reduce all the constant terms (Y's) by a factor C, and multiply the results by the same factor.

Should circuit instability be present (oscillation) with the determinant set up, the cause is due to the presence of one or more characteristic roots with negative real parts. Before proceeding further, this unstable condition must be remedied. This may be done by interchanging rows or columns, changing the sign of one or more equations, or by more drastic means (which will always work) such as multiplying the matrix by its transpose to obtain a positive definite determinant. Normally, the equations should be arranged so that the diagonal coefficients are the largest ones (if possible) and have positive signs. This does not necessarily insure that all roots will have positive real parts.

For systems of equations of five or less, stable operation may be insured by setting the matrix and its negative transpose in the solver as follows. The system of three equations will be used for the example.

Constants				Coefficients		
0	1	0	0	$-a_{1-1}$	$-a_{2-1}$	$-a_{3-1}$
0	0	1	0	$-a_{1-2}$	$-a_{2-2}$	$-a_{3-2}$
0	0	0	1	$-a_{1-3}$	$-a_{2-3}$	$-a_{3-3}$.
Y_1	a_{1-1}	a_{1-2}	a_{1-3}	0	0	0
Y_2	a_{2-1}	a_{2-2}	a_{2-3}	0	0	0
Y_3	a_{3-1}	a_{3-2}	a_{3-3}	0	0	0

(21)

The results would then be measured in the normal manner. This system will always be stable, but the precaution regarding prevention of amplifier overloading should be observed.

Accuracy obtained depends upon the nature of the system of equations, as well as the characteristics of the equation solver. Should all the characteristic roots of the determinant be in the neighborhood of unity or larger, the results obtained would depend mainly upon the components in the equation solver, in which case the results would satisfy the individual equations within about one-tenth percent. If the determinant has some small characteristic roots, some of the results will be poorly defined. This is not the fault of the equation solver, but is an inherent characteristic of the system of equations. Such cases where the results are poorly defined indicate that physically, values of the poorly defined unknowns may vary considerably, yet satisfy the equations fairly well.

Essentially unlimited accuracy may be obtained if an iteration process is employed. The process is as follows:

1. Set up equations and solve for unknowns (X's) in the normal manner.

2. Substitute the values obtained for the X's into the original equations. Take the difference between the values thus obtained and the values the equation should have (Y's). Call these differences ΔY_1, ΔY_2, etc.

3. Change the Y dials only to $K\Delta Y_1$, $K\Delta Y_2$, etc. (K generally being 100 or 1000 and chosen to utilize as much scale factor as possible) and measure $K\Delta X_1$, $K\Delta X_2$, etc.

4. Add ΔX_1, ΔX_2, etc., to X_1, X_2, etc., to obtain corrected values for the answers. If still higher accuracy is desired, repeat the process.

The following example should clarify the process described above:

Equation to be solved:

$$\begin{aligned} &\text{I.}\quad 1 = 1X_1 + 0.4X_2 + 0.1X_3, \\ &\text{II.}\quad 0 = -0.9X_1 + 1X_2 + 0.4X_3, \\ &\text{III.}\quad 0 = -0.2X_1 - 0.4X_2 + 1X_3. \end{aligned} \tag{22}$$

Direct solution obtained from equation solver. (Superscripts denote number of iterations which have taken place.)

$$X_1^0 = +0.753, \quad X_2^0 = +0.531, \quad X_3^0 = +0.363. \tag{23}$$

Checking the equations by substitution

I	II	III
$1X_1 = +0.7530$	$-0.9X_1 = -0.6777$	$-0.2X_1 = -0.1506$
$0.4X_2 = +0.2124$	$1X_2 = +0.5310$	$-0.4X_2 = -0.2124$
$0.1X_3 = +0.0363$	$0.4X_3 = +0.1452$	$1X_3 = +0.3630$
$Y_1^0 = +1.0017$	$Y_2^0 = -0.0015$	$Y_3^0 = 0.0000$
$Y_1 = 1.0000$	$Y_2 = 0.0000$	$Y_3 = 0.0000$
$\Delta Y_1^0 = -0.0017$	$\Delta Y_2^0 = +0.0015$	$\Delta Y_3^0 = 0.0000$

(24)

(25)

Let $K = 100$, and set

$$K\Delta Y_1 = -0.17, \quad K\Delta Y_2 = +0.15, \quad K\Delta Y_3 = 0.00, \tag{26}$$

in place of the original Y's in the equation solver, and solve for the corrections to be added to the first X's obtained. The values obtained for the correction terms are

$$\Delta X_1' = -0.00171, \quad \Delta X_2' = +0.00007, \quad \Delta X_3' = -0.00031, \tag{27}$$

and the corrected values become

$X_1^0 = +0.75300$	$X_2^0 = +0.53100$	$X_3^0 = +0.36300$
$\Delta X_1' = -0.00171$	$\Delta X_2' = +0.00007$	$\Delta X_3' = -0.00031$
$X_1' = +0.75129$	$X_2' = +0.53107$	$X_3' = 0.36269$

(28)

The corrected values for the X's may now be checked.

I	II	III
$X_1' = +0.751290$	$-0.9X_1' = -0.676161$	$-0.2X_1 = -0.150258$
$0.4X_2' = +0.212428$	$1X_2' = +0.531070$	$-0.4X_2 = -0.212428$
$0.1X_3' = +0.036269$	$+0.4X_3 = +0.145076$	$+1X_3 = +0.362690$
$Y_1' = 0.999987$	$Y_2' = -0.000015$	$Y_3' = 0.000004$

(29)

The errors in the results after one iteration for the above system of equations are

$$\begin{aligned} \Delta Y_1' &= 1.000000 - 0.999987 = +0.000013, \\ \Delta Y_2' &= 0.000000 - (-0.000015) = +0.000015, \\ \Delta Y_3' &= 0.000000 - 0.000004 = -0.000004. \end{aligned} \tag{30}$$

The iteration process could be repeated again if still higher accuracy were desired. For most engineering problems, the direct answers (superscript 0) are normally accurate enough. The iteration process goes very rapidly when higher accuracy is desired, especially if a standard desk calculator is available.

Higher direct accuracy could be obtained if a bridging system and an accurate four-dial decade were used for adjusting the values of the coefficients. It is felt that this is not worth while, however, since iteration goes very rapidly, and much better accuracy is obtained after only one iteration than could be obtained with the bridging system for setting coefficients, and also since it would take a considerably longer time to set up the equations on the solver with the bridging system rather than with the direct reading dials.

The results are always present in the form of

voltages. A bank of meters could be employed to give a continuous indication of all the answers simultaneously. This feature would be very useful in certain types of problems where it is desirable to note the effect of varying certain coefficients and constants on all the results.

Simultaneous equations with complex constant and coefficient terms may be solved by breaking the system up into a system of twice as many real equations by equating real and imaginary parts separately. For example, the following two equations with complex parts may be solved by converting them to a system of four equations.

$$\begin{aligned} U_1+jV_1&=(a_{1-1}+jb_{1-1})(X_1+jZ_1)\\ &\quad+(a_{1-2}+jb_{1-2})(X_2+jZ_2),\\ U_2+jV_2&=(a_{2-1}+jb_{2-1})(X_1+jZ_1)\\ &\quad+(a_{2-2}+jb_{2-2})(X_2+jZ_2). \end{aligned} \tag{31}$$

The above equations become

$$\begin{aligned} U_1&=a_{1-1}X_1-b_{1-1}Z_1+a_{1-2}X_2-b_{1-2}Z_2,\\ V_1&=b_{1-1}X_1+a_{1-1}Z_1+b_{1-2}X_2+a_{1-2}Z_2,\\ U_2&=a_{2-1}X_1-b_{2-1}Z_1+a_{2-2}X_2-b_{2-2}Z_2,\\ V_2&=b_{2-1}X_1+a_{2-1}Z_1+b_{2-2}X_2+a_{2-2}Z_2. \end{aligned} \tag{32}$$

A.C. circuit problems are of this general nature. Problems involving linear passive networks will always result in a system of equations which will be stable on the equation solver if the original equations with the complex coefficients are written in the mesh order mentioned in the section on stability.

CIRCUIT AND COMPONENT FEATURES

Each of the components in the main network is of ±0.1 percent accuracy. The summing resistors are 500,000 ohm ±0.1 percent wire wound, and the potentiometers are 1000-ohm micropots with a guaranteed linearity of ±0.1 percent. Positive and negative values for coefficients and constants are provided for by making all amplifier output circuits push-pull. This is conveniently done by utilizing output stages with equal impedances in the plate and cathode circuits. Figure 2 illustrates this circuit. Each micropot is accompanied by a matching resistor of equal total value. A four-pole, double-throw switch is used for either (a) inserting the micro-

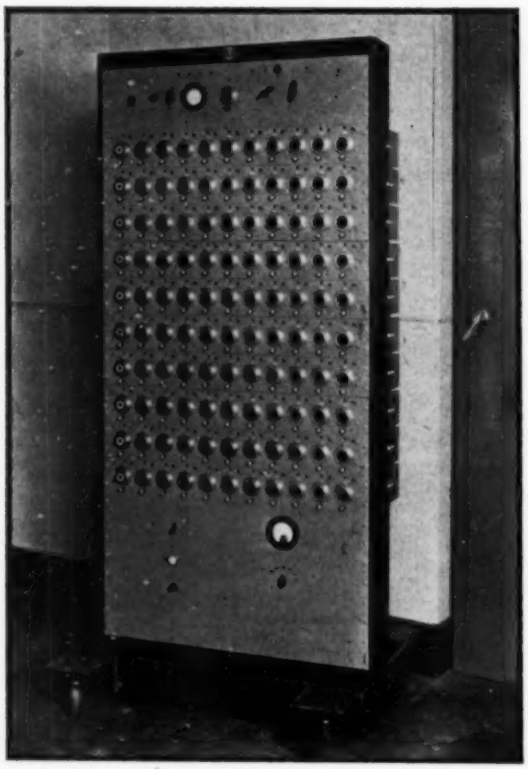


FIG. 3. Front view of a ten equation computer.

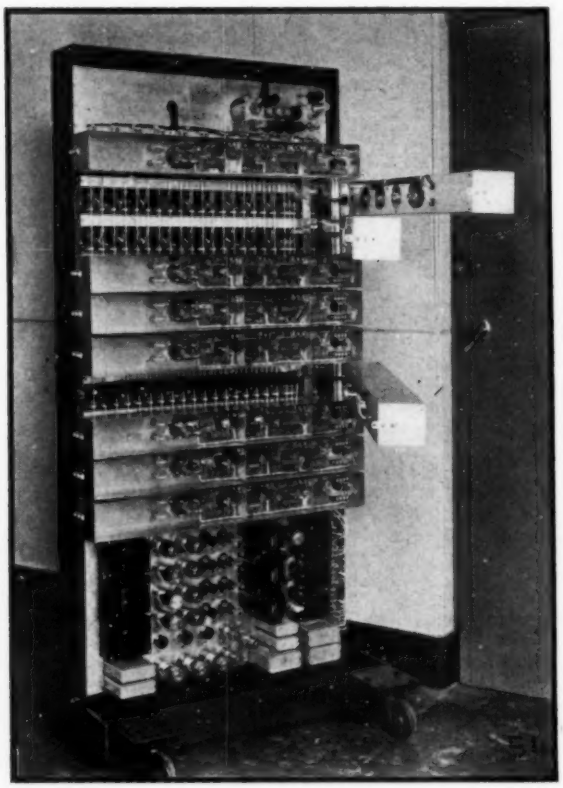


FIG. 4. Rear view of a ten equation computer.

pot in the cathode circuit and the matching resistor in the plate circuit (positive coefficient or constant), or (b) inserting the micropot in the plate circuit and the matching resistor in the cathode circuit (negative coefficient or constant). This combination exists for each coefficient, constant, and the bridging circuit for measuring the X's. There are 111 such combinations in the equation solver for ten equations.

In order to utilize the system just described for making available accurately balanced push-pull signals, a very highly regulated power supply (extremely low internal impedance) was required. The unit incorporated has a dynamic impedance at the operating frequency (1000 cycles) of about 0.0001 ohm. For other frequencies, it is somewhat greater, but is still extremely low over an extremely wide range.

Compensation circuits to correct for the fact that most micropots cannot be set for zero output, and do not have the same electrical travel for 3600° mechanical travel were incorporated. Since the corrections required for these items were small in all cases, carbon resistors were used in the compensation networks.

The networks for each equation are assembled on different panels which may be individually removed. Each amplifier is mounted in an individual hinged shield can so it may be removed without disturbing anything else. All units are thus readily accessible for easy servicing in case a failure of any kind develops. Presence of possible trouble is readily detected upon checking the answers to a problem. Trouble can be very easily located by a very simple procedure, and normally remedied very quickly. Failures which have occurred with the laboratory model have been mainly a result of faulty tubes. These instances have been few.

Figure 3 is a photograph of the front side of the equation solver. The top panel contains (center—left to right) the following: signal amplitude control, vertical gain control, horizontal gain control, cathode-ray tube bridge balance indicator, unknown (X) measuring control, unknown (X) selector switch, and unknown (X) multiplier. Above these controls are located screwdriver adjustments for centering, brightness, focus, and bridge phasing. The next ten panels contain the controls for setting the constant (Y's) and coefficient (a's) terms of each equation. A power switch, fuse, voltmeter, and switch for connecting the voltmeter to any one of the power supplies occupy the lowest panel. Dials representing the constant terms (Y's) have white circles painted on them, while those representing the diagonal coefficients (a_{1-1}, a_{2-2}, a_{3-3}, etc.) have red circles. The lever switch under each dial is used for setting the sign of the constant or coefficient term.

Figure 4, a rear view, illustrates the disposition of the signal generator and bridging circuits, amplifiers, and power supplies. Three of the amplifiers are swung out to expose the rack wiring.

ACKNOWLEDGMENT

The authors wish to express their appreciation for the work done by J. M. Morgan on the mechanical design of the equipment, and for the encouragement given by Dr. V. K. Zworykin in the execution of the project.

Surface Waves from an Underwater Explosion

J. G. KIRKWOOD,* *Cornell University, Ithaca, New York*

AND

R. J. SEEGER,** *Bureau of Ordnance, Washington, D. C.*

(Received August 8, 1947)

A general expression has been derived theoretically for the pressure due to surface waves produced by an underwater explosion. The integrals involved have been evaluated in terms of dimensionless variables over a range of practical interest. A specific formula is given for the pressure on the bottom of the water due to the first pulsation of the gas globe from a TNT explosion originating also on the bottom.

CONSIDER a surface wave generated by an underwater explosion.[1] The effect of the shock wave is assumed to be negligible so that the sole source is the pulsating gas globe, which has a volume $V(t)$ at the time t. This gas globe, assumed to be spherically symmetric, as a first approximation, will migrate under the influence of gravity and of the neighboring surfaces. Let $z_B(t)$ (compare Fig. 1) designate its position above the bottom. It is required to find the flow for infinite water of uniform depth h above a rigid bottom.

Let P be a point (Fig. 1) with cylindrical polar coordinates r, θ, z. Then the conservation of mass requires that the divergence of the material velocity $\mathbf{q}(r, \theta, z, t)$ of an inviscid, incompressible fluid with uniform density ρ, vanish everywhere except at the source (equation of continuity), i.e.,

$$\nabla \cdot \mathbf{q} = 0.$$

If the flow is further assumed to be irrotational, then the velocity, $\mathbf{q}$, is the negative gradient of a velocity potential, $\Phi(r, \theta, z, t)$, and Laplace's equation results from the above one, i.e.,

$$\nabla^2 \Phi = 0. \tag{1}$$

In addition, the function Φ must satisfy two boundary conditions: At the fixed bottom there is no component of the velocity normal to the surface, i.e.,

$$\partial \Phi / \partial z = 0, \quad z = 0. \tag{2}$$

At the free surface the pressure p_a is uniformly atmospheric. The effect of this condition is seen by a consideration of the Eulerian equation of motion, *viz.*,

$$(\partial \mathbf{q}/\partial t) + \mathbf{q} \cdot \nabla \mathbf{q} = -\nabla(gz) - (1/\rho)\nabla p,$$

TABLE I.

h (ft.) \ W (lb.)	300 L	1200 L	2100 L	6600 L
25	24.21	—		
30	23.55	—	—	—
35	22.96	36.45	—	—
40	22.43	35.60	—	—
45	21.94	34.82	41.96	—
50	21.49	34.11	41.10	—
55	21.07	33.45	40.31	59.04
60	20.68	32.84	39.57	57.96
70	19.99	31.74	38.25	56.02
80	19.39	30.77	37.08	54.32
90	18.84	29.92	36.05	52.81
100	18.36	29.15	35.12	51.45
110	17.92	28.45	34.28	50.22
120	17.52	27.82	33.52	49.10

TABLE II.

h (ft.) \ W (lb.)	300 τ	300 τ'	1200 τ	1200 τ'	2100 τ	2100 τ'	6600 τ	6600 τ'
25	0.8741	0.9920						
30	0.8575	0.8883		—		—		—
35	0.8330	0.7990	1.192	1.143		—		—
40	0.8065	0.7236	1.170	1.050				
45	0.7790	0.6590	1.143	0.9666	1.319	1.116		—
50	0.7522	0.6037	1.112	0.8926	1.292	1.037		—
55	0.7262	0.5557	1.081	0.8269	1.259	0.9632	—	—
60	0.7018	0.5141	1.050	0.7694	1.226	0.8984	1.653	1.211
70	0.6571	0.4457	0.9909	0.6721	1.163	0.7885	1.589	1.073
80	0.6168	0.3913	0.9356	0.5936	1.102	0.6991	1.519	0.9639
90	0.5810	0.3475	0.8859	0.5299	1.046	0.6254	1.452	0.8685
100	0.5492	0.3116	0.8414	0.4774	0.9943	0.5642	1.388	0.7878
110	0.5209	0.2818	0.8005	0.4331	0.9479	0.5129	1.329	0.7193
120	0.4954	0.2566	0.7632	0.3953	0.9059	0.4693	1.274	0.6597

* Now at the California Institute of Technology.

** Now at the Naval Ordnance Laboratory, White Oak, Maryland.

[1] The theoretical treatment is an application of the standard procedure for surface waves (see H. Lamb's *Hydrodynamics*, 6th edition, Chapter IX). It has already been briefly reported in J. G. Kirkwood's "Memorandum on the generation of surface waves by an underwater explosion" (Undex 94). The present report embodies also certain suggestions made by R. J. Finkelstein, J. von Neumann, and F. J. Weyl of the Navy Bureau of Ordnance Research Group on the theory of explosives.

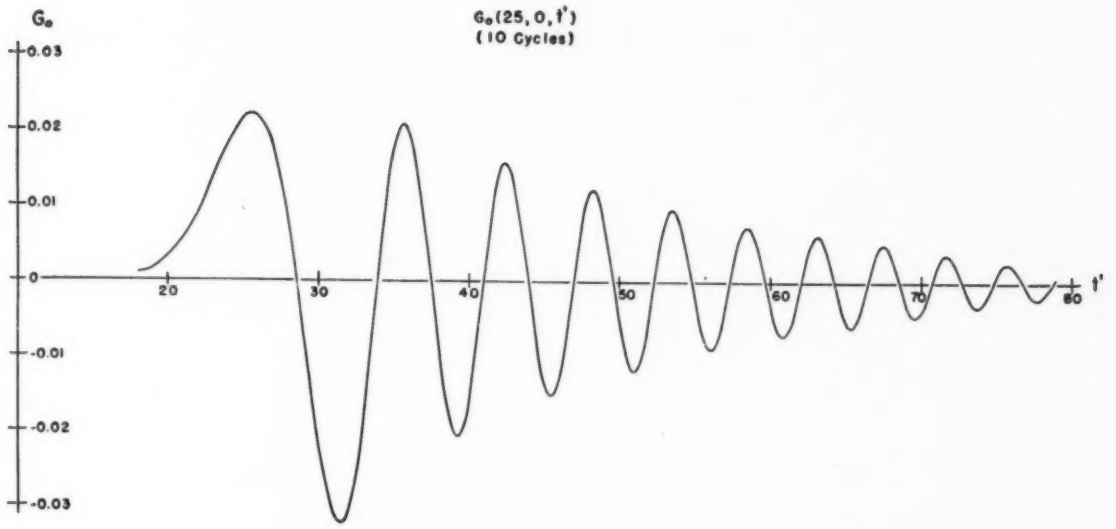


FIG. 1a.

where (gz) is the gravitational potential and $p(r, \theta, z, t)$ is the fluid pressure at a point. Integrating with respect to the coordinates and neglecting the square of the speed $\mathbf{q}$, we obtain

$$-(\partial\Phi/\partial t)+gz+(p/\rho)=0, \qquad (3a)$$

where an arbitrary integration function of t has been absorbed in $\partial\Phi/\partial t$. Including in it also the constant p_a/ρ on the surface, we have, to the same order of approximation, for a small disturbance of the free surface

$$-(\partial\Phi/\partial t)_{z=h}+gz=0, \quad \text{free surface.}$$

Now kinematically the velocity of the fluid normal to the surface must be the same as the normal component of the surface velocity so that approximately (to first-order terms)

$$-(\partial\Phi/\partial z)_{z=h}=(\partial z/\partial t), \quad \text{free surface.}$$

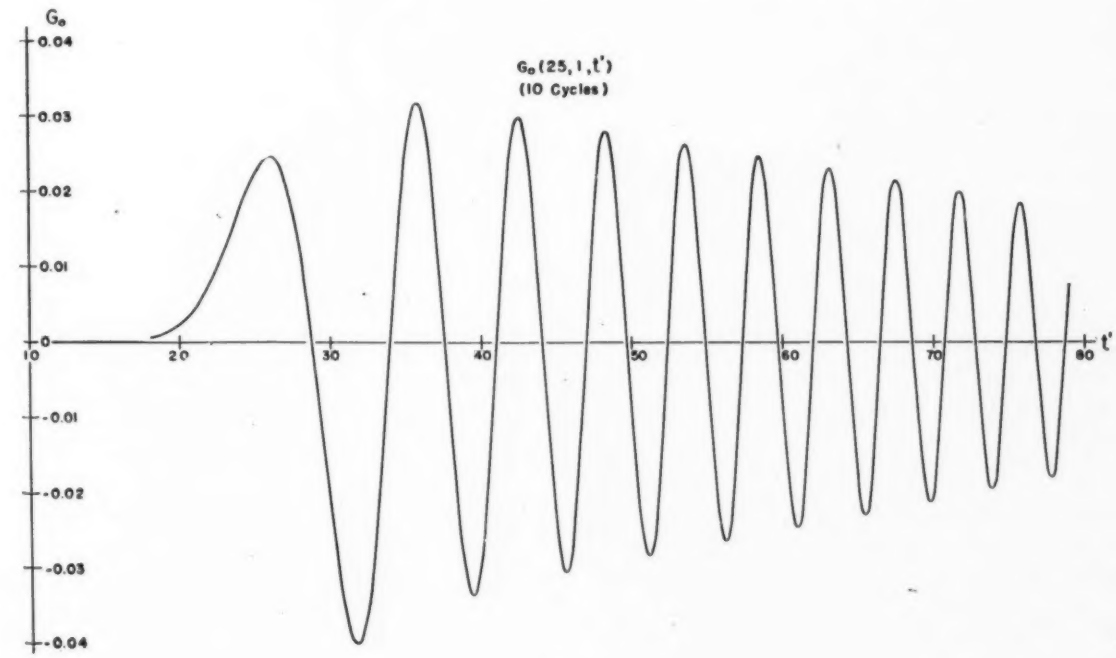


FIG. 1b.

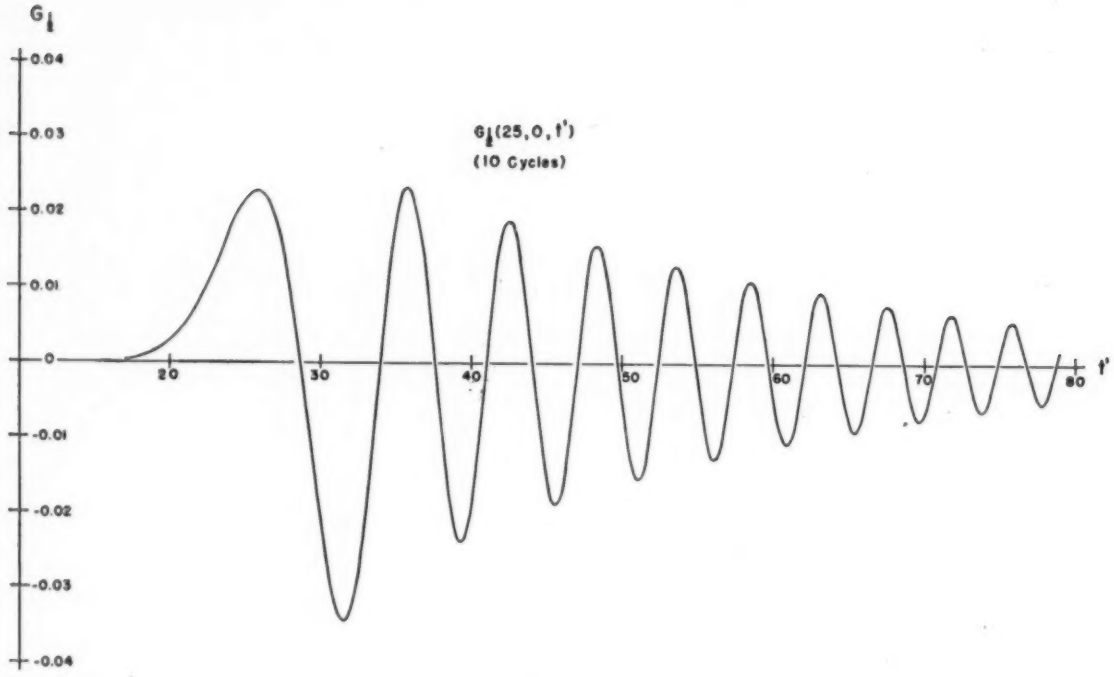


FIG. 1c.

Differentiating the preceding equation with respect to t and substituting for $\partial z/\partial t$, we obtain for our second boundary condition

$$(\partial^2\Phi/\partial t^2)+g(\partial\Phi/\partial z)=0, \quad z=h. \tag{3}$$

Suppose $\Phi_B(r, z, t)$ is the cylindrically symmetric velocity potential for the case of a bottom only, i.e., no free surface. Then

$$\Phi_B=\frac{1}{4\pi}\frac{dV}{dt}\left\{\frac{1}{(r^2+(z-z_E)^2)^{\frac{1}{2}}}+\frac{1}{(r^2+(z+z_E)^2)^{\frac{1}{2}}}\right\}, \tag{4a}$$

where dV/dt is the strength of an assumed simple source, i.e., a point source with uniform radial flow. Let $\Phi_s(r, z, t)$ be the additional term required for a free surface to be satisfied also. Thus,

$$\Phi=\Phi_B+\Phi_s. \tag{4}$$

Inasmuch as Φ_B satisfies Laplace's equation subject to the condition of the fixed bottom, it follows from Eqs. (1) and (2), respectively, that

$$\nabla^2\Phi_s=0, \tag{1'}$$

and

$$\partial\Phi_s/\partial z=0, \quad z=0. \tag{2'}$$

Substituting the expression (4) for Φ in Eq. (3), we have

$$\frac{\partial^2\Phi_s}{\partial t^2}+g\frac{\partial\Phi_s}{\partial z}=-\frac{\partial^2\Phi_B}{\partial t^2}-g\frac{\partial\Phi_B}{\partial z}, \quad z=h,$$

or, upon substituting for Φ_B the expression (4a),

$$\left(\frac{\partial^2\Phi}{\partial t^2}\right)_{z=h}+g\left(\frac{\partial\Phi_s}{\partial z}\right)_{z=h}=-\left(\frac{\partial^2\Phi_B}{\partial t^2}\right)_{z=h}+\frac{g}{4\pi}\frac{dV}{dt}\left\{\frac{h-z_E}{(r^2+(h-z_E)^2)^{\frac{1}{2}}}+\frac{h+z_E}{(r^2+(h+z_E)^2)^{\frac{1}{2}}}\right\}. \tag{3'}$$

We shall assume a Hankel transform $f(k, z, t)$ for the cylindrically symmetric solution $\Phi_s(r, z, t)$,

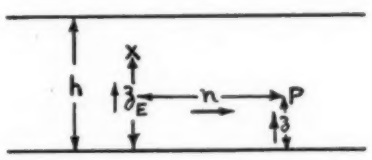


FIG. 1d.

FIG. 2a.

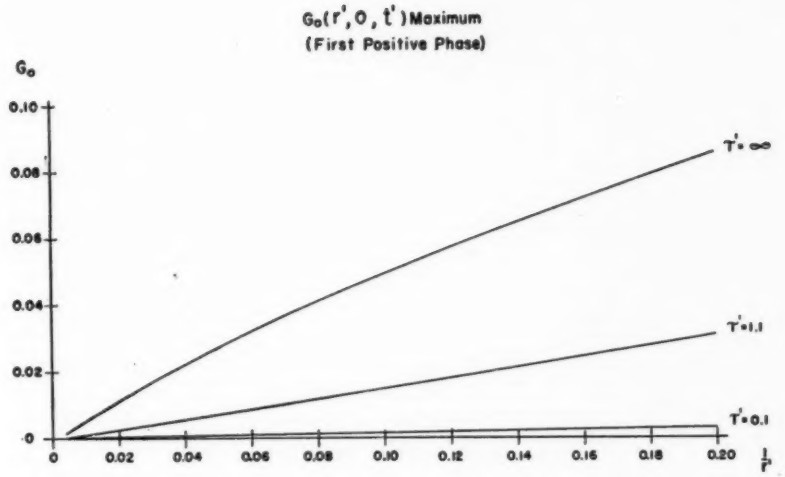
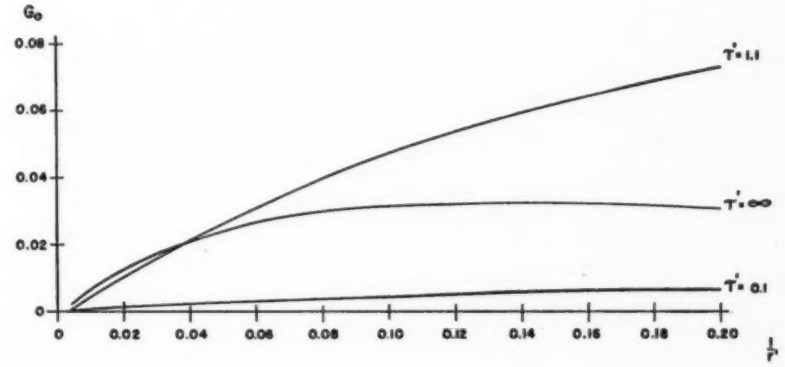


FIG. 2b.



namely,

$$\Phi_s(r, z, t) = {}_rT_k(f(k, z, t)), \tag{4b}$$

where

$${}_rT_k(f(k, z, t)) \equiv \int_0^\infty f(k, z, t) J_0(kr) k dk,$$

and $J_0(kr)$ is the Bessel function of zero order. Substituting this expression for Φ_s in Eqs. (1′) and (2′) and commuting operators, we obtain, respectively,

$${}_rT_k\left(\frac{\partial^2 f}{\partial t^2} - k^2 f\right) = 0,$$

$${}_rT_k\left(\frac{\partial f}{\partial z}\right)_{z=0} = 0.$$

Applying the inverse transformation and recalling that iteration of this operator is equivalent to multiplication by unity, we find

$$(\partial^2 f/\partial z^2) - k^2 f = 0,$$
$$(\partial f/\partial z)_{z=0} = 0.$$

These conditions are satisfied by

$$f(k, z, t) = K(k, t) \cosh kz.$$

Substituting Φ_s from (4b) also in Eq. (3′) and commuting operators, we have

$${}_rT_k\left(\frac{\partial^2 f}{\partial t^2} + g\frac{\partial f}{\partial z}\right)_{z=h} = -\left(\frac{\partial^2 \Phi_B}{\partial t^2}\right)_{z=h} + \frac{g}{4\pi}\frac{dV}{dt}\left\{\frac{h - z_E}{(r^2 + (h - z_E)^2)^{\frac{1}{2}}} + \frac{h + z_E}{(r^2 + (h + z_E)^2)^{\frac{1}{2}}}\right\}. \tag{3′}$$

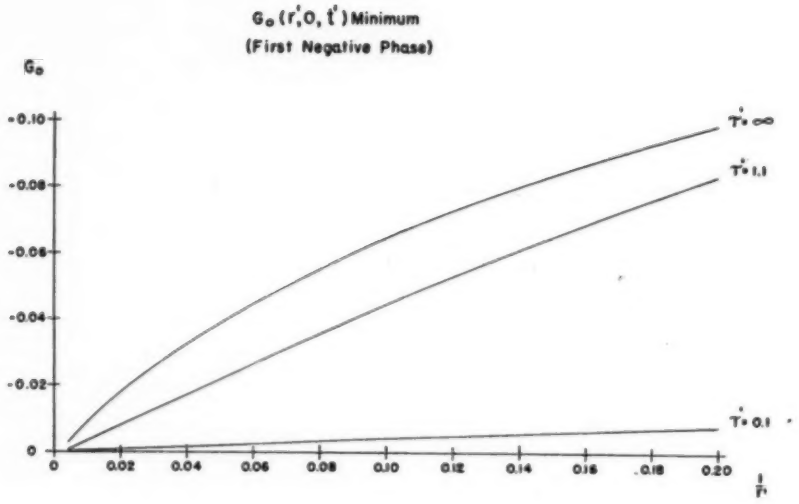


FIG. 2c.

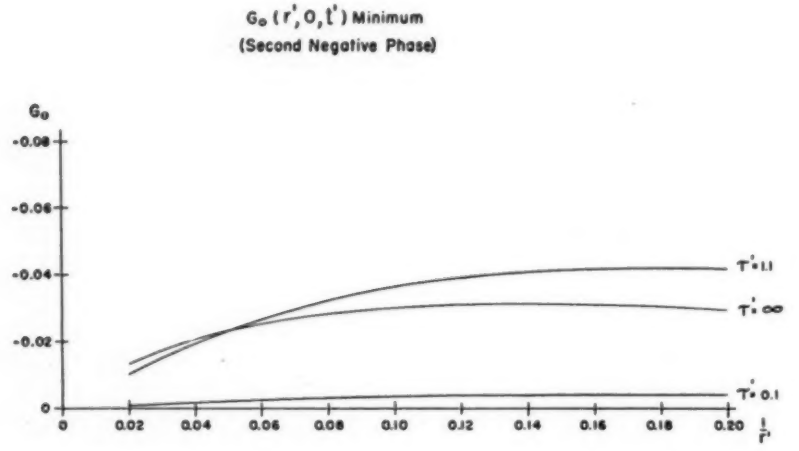


FIG. 2d.

Applying the inverse and noting that

$$ {}_kT_r\frac{u}{(r^2+u^2)^{\frac{1}{2}}}=e^{-ku}, \quad u>0, $$

we obtain

$$ \frac{\partial^2 K}{\partial t^2}\cosh kh+gkK\sinh kh=-{}_kT_r\left(\frac{\partial^2\Phi_B}{\partial t^2}\right)_{z=h} $$

$$ +\frac{g}{4\pi}\frac{dV}{dt}(e^{-k(h-z_E)}+e^{-k(h+z_E)}). $$

Put

$$ \omega\equiv(gk\tanh kh)^{\frac{1}{2}}, $$

then

$$ \frac{\partial^2 K}{\partial t^2}+\omega^2 K=-\frac{1}{\cosh kh}\,{}_kT_r\left(\frac{\partial^2\Phi_B}{\partial t^2}\right)_{z=h} $$

$$ +\frac{g\exp(-kh)}{2\pi\cosh kh}\cosh kz_E. $$

The real solution of this equation is given by

$$ K(k,t)=A(k)\cos(\omega t-\alpha(k)) $$

$$ +\frac{1}{\omega\cosh kh}\int_{t_0}^{t}\left\{-{}_kT_r\left(\frac{\partial^2\Phi_B}{\partial s^2}\right)_{z=h}\right. $$

$$ \left.+\frac{g}{2\pi}e^{-kh}\frac{dV}{ds}\cosh kz_E\right\}\sin\omega(t-s)ds, $$

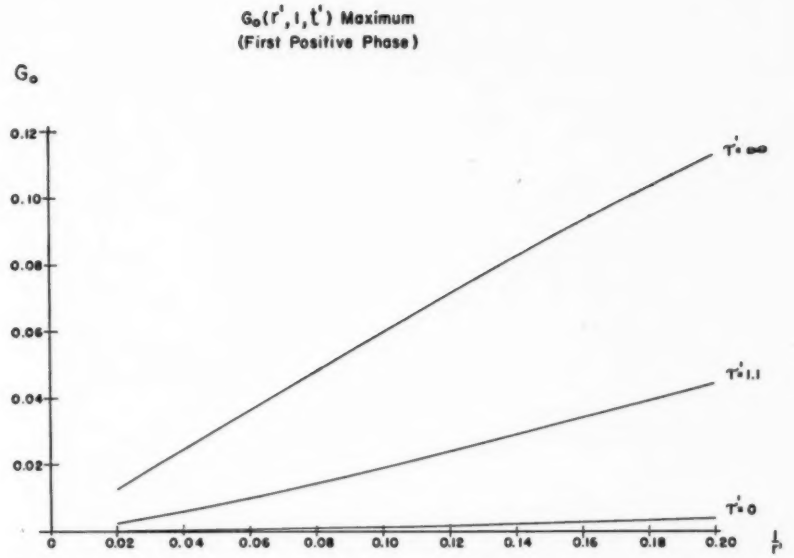


FIG. 2e.

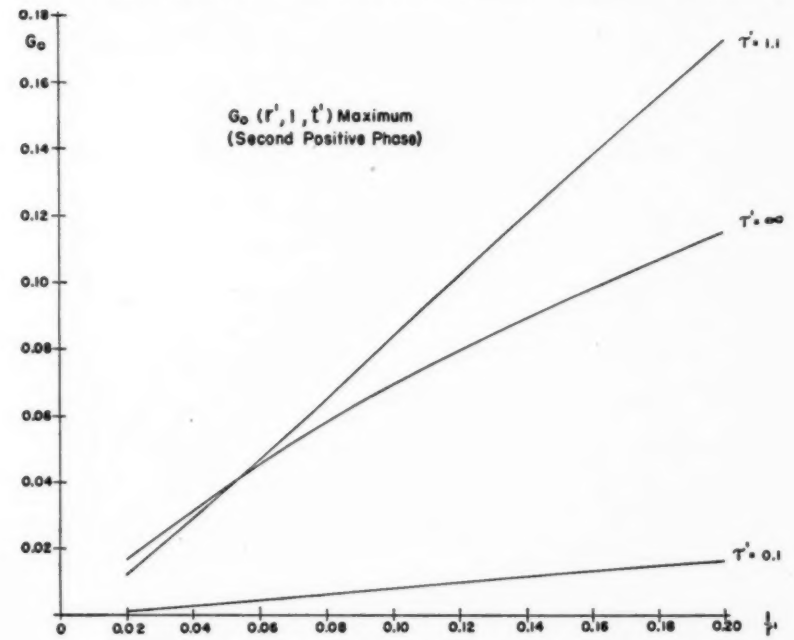


Fig. 2f.

where t_0 is some time prior to the explosion and $A(k)$, $\alpha(k)$ are arbitrary constants of integration. Let the explosion take place at the time $t=0$. For all times previous to this instant $\Phi_s(r, z, t)$, and consequently both $K(k, t)$ and $A(k)$, must vanish. Therefore,

$$K(k,t)=\frac{1}{\omega\cosh kh}\int_{t_0}^{t}\left\{-{}_kT_r\left(\frac{\partial^2\Phi_B}{\partial s^2}\right)_{z=h}+\frac{g}{2\pi}e^{-kh}\frac{dV}{dt}\cosh kz_E\right\}ds.$$

If the gas globe is stationary, i.e., z_E constant, then from the expression (4a) for Φ_B

$${}_kT_r\left(\frac{\partial^2\Phi_B}{\partial s^2}\right)_{z=h}=\frac{1}{4\pi}\frac{d^3V}{ds^3}\,{}_kT_r\times\left\{\frac{1}{(r^2+(h-z_E)^2)^{\frac{1}{2}}}+\frac{1}{(r^2+(h+z_E)^2)^{\frac{1}{2}}}\right\}.$$

Noting that

$${}_kT_r\frac{1}{(r^2+u^2)^{\frac{1}{2}}}=\frac{\exp(-ku)}{k},\quad u>0,$$

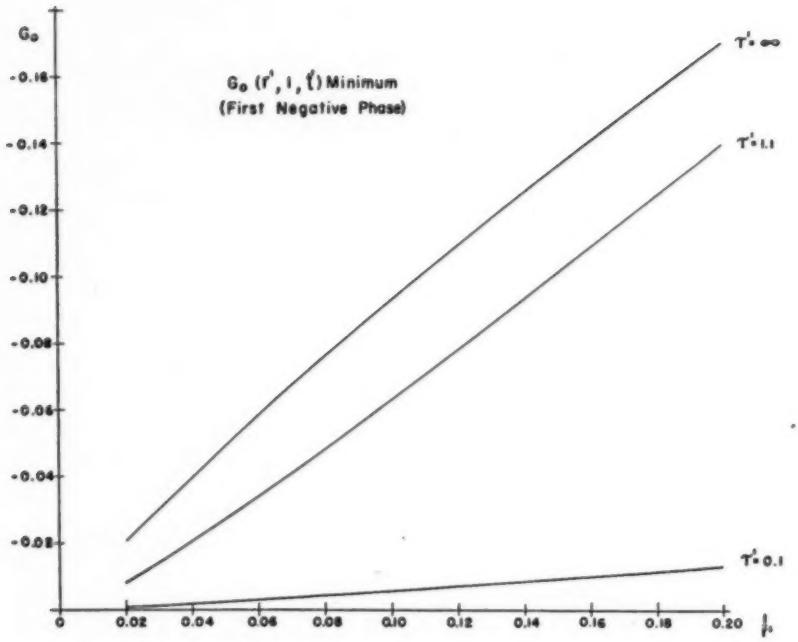


FIG. 2g.

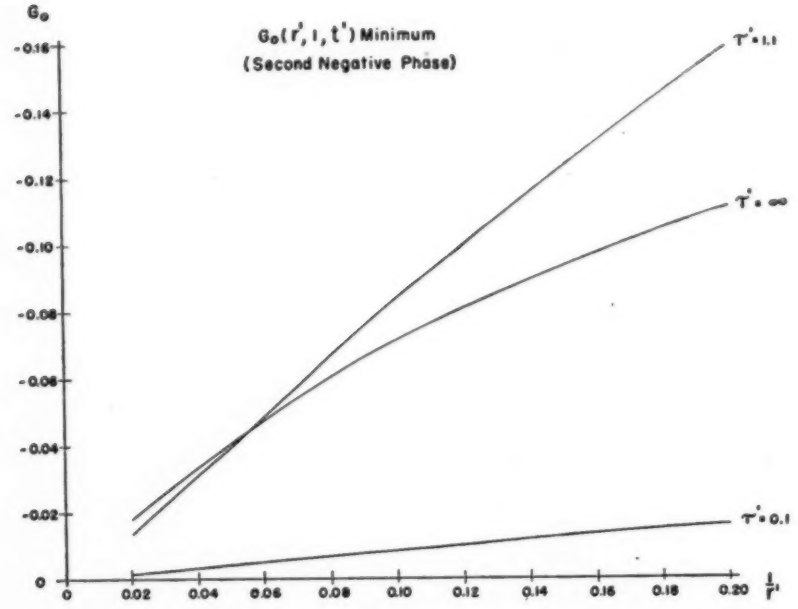


FIG. 2h.

we obtain

$$K(k,t)=\frac{\exp(-kh)\cosh kz_E}{2\pi\omega\cosh kh}\times\int_{t_0}^{t}\left(-\frac{1}{k}\frac{d^3V}{ds^3}+g\frac{dV}{ds}\right)\sin\omega(t-s)ds.$$

Integrating by parts, we find

$$K(k,t)=\frac{\cosh kz_E}{2\pi\cosh kh}\left\{-\frac{\exp(-kh)}{k}\frac{dV}{dt}+\frac{g}{\cosh kh}\int_{t_0}^{t}V(s)\cos\omega(t-s)ds\right\}.$$

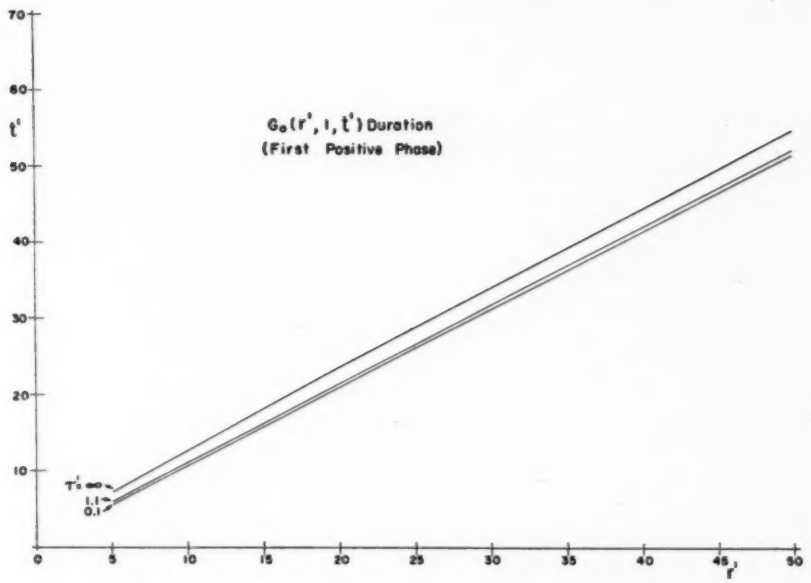
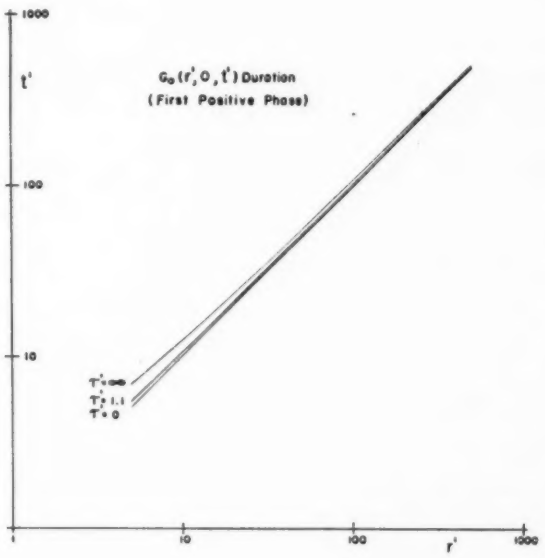


FIG. 3a.



Hence, the complete solution Φ is given by

$$\Phi(r, z, t) = \Phi_B(r, z, t) + \frac{1}{2\pi} {}_rT_k \times \left\{ \frac{\cosh kz \cosh kz_E}{\cosh kh} \left(-\frac{\exp(-kh)}{k} \frac{dV}{dt} + \frac{g}{\cosh kh} \int_{t_0}^{t} V(s) \cos\omega(t-s) ds \right) \right\}. \quad (4')$$

The evaluation of the integrals involved in Φ for an actual gas globe is straightforward, but lengthy. It is convenient to introduce, therefore, a simplifying assumption, the value of which must be tested by analysis of the experimental data. If the period, τ, of the first pulsation of the gas globe is much less than the time interval after the explosion, it is reasonable to suppose that

$$V(t) = \bar{V}, \quad 0 \leqslant t < \tau,$$

where the constant $\bar{V}$ is some average volume for

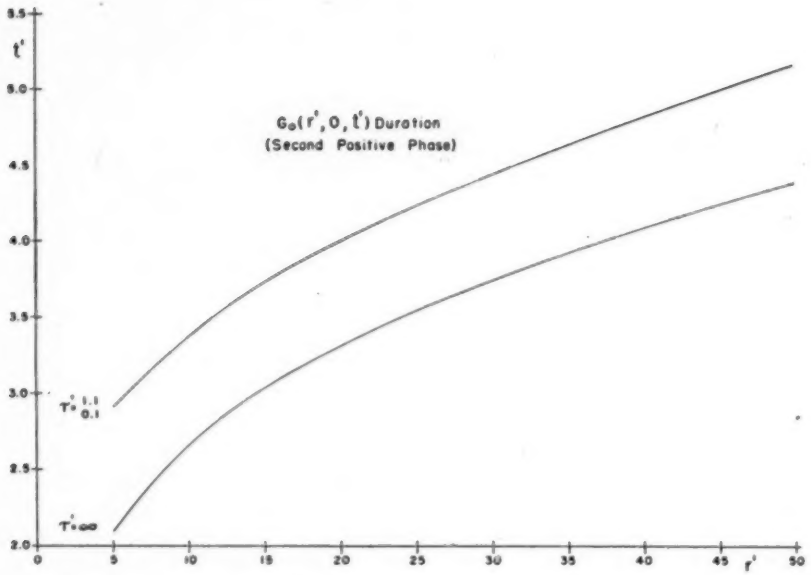


FIG. 3b.

the period τ. Thus

$$\Phi(r, z, t) = \frac{g}{2\pi}\bar{V}_{,T_k}\frac{\cosh kz \cosh kz_E}{\omega \cosh^2 kh} \times \{\sin\omega t - \sin\omega(t-\tau)\}, \quad (4'')$$

and from Eq. (3a)

$$p(r, z, t) = -\rho gz + \frac{\rho g}{2\pi}\bar{V}_{,T_k} \times \frac{\cosh kz \cosh kz_E}{\cosh^2 kh}\{\cos\omega t - \cos\omega(t-\tau)\}. \quad (5)$$

For computational purposes it is convenient to introduce dimensionless variables, namely,

$$r' \equiv r/h, \quad z' \equiv z/h, \quad z_E' \equiv z_E/h,$$

and

$$t' \equiv t(g/h)^{\frac{1}{2}}, \quad \tau' \equiv \tau(g/h)^{\frac{1}{2}}, \quad \omega' \equiv \omega(h/g)^{\frac{1}{2}}.$$

Also put

$$\beta = kh,$$

then

$$\omega' = (\beta \tanh\beta)^{\frac{1}{2}}.$$

Hence,

$$p_{z'E}(r', z', t') = -\rho ghz' + (\rho g/2\pi h^2)\bar{V}G_{z'E,\tau'}(r', g', t'), \quad (5')$$

where

$$G_{z'E,\tau'}(r', z', t') \equiv G_{z'E}(r', z', t') - G_{z'E}(r', z', t'-\tau'),$$

and

$$G_{z'E}(r', z', t') \equiv \int_0^\infty \frac{\cosh\beta z' \cosh\beta z_E'}{\cosh^2\beta} \cos\omega' t' \cdot J_0(\beta r')\beta d\beta.$$

A problem of practical interest is the determination of the pressure due to an explosion at the bottom ($z_E' = 0$). The function $G_0(r', 0, t')$ has been evaluated[2] numerically for $r' = 5, 10, 15, 25, 50, 500$; the function $G_0(r', 1, t')$ [equal to $G_1(r', 0, t')$] has been evaluated for $r' = 5, 10, 15, 25, 50$.

The function $G_{z'E}(r', 0, t')$ has apparently alternately positive and negative phases, as is seen in typical representations, $G_0(25, 0, t')$, $G_0(25, 1, t')$, $G_{\frac{1}{2}}(25, 0, t')$ in Figs. 1a, b, and c, respectively. The maximum of $G_0(r', 0, t')$ for the first positive phase and that for the second one are shown for the several values of r' in Figs. 2a, 2b, respectively; likewise, the minima of the first two negative phases are given for corresponding values of r' in Figs. 2c, 2d. The durations of

[2] These integrals have all been evaluated by the Mathematics Tables Project under the Applied Mathematics Panel of the NDRC; they are published in NavOrd Report No. 401. The authors are particularly indebted to G. Blanch and A. N. Lowan of that group.

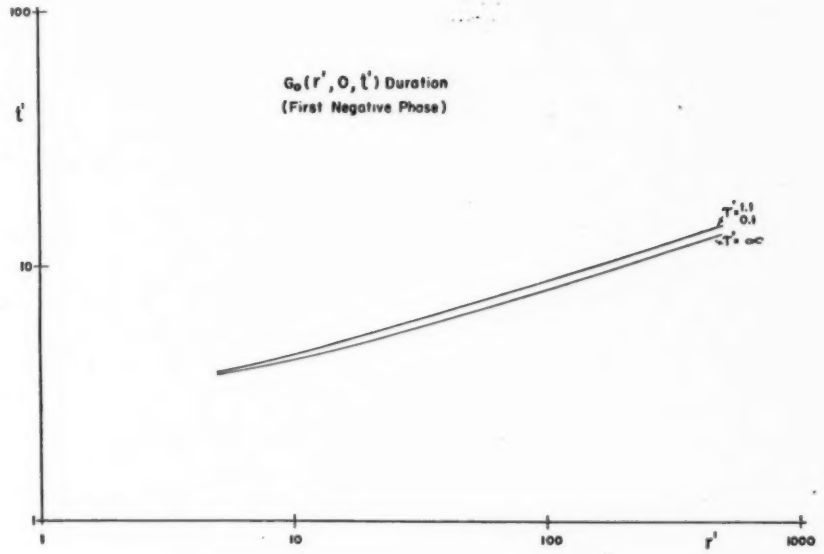
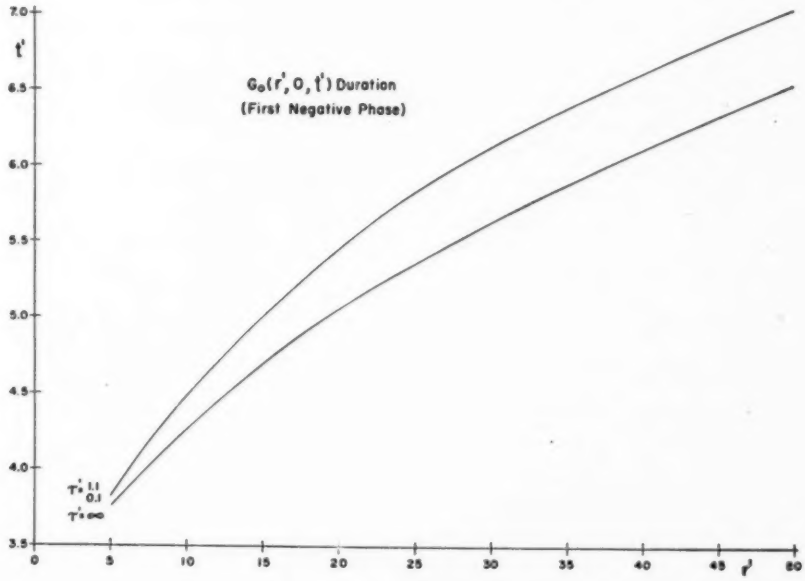


FIG. 3c.



the corresponding phases are represented in Figs. 3a, 3b (positive) and in Figs. 3c, 3d (negative). The corresponding values for $G_0(r', 1, t')$ are plotted in Figs. 2e, 2f, 2g, 2h, 3e, 3f, 3g, 3h. It is to be noted from Fig. 4 that the maximum of the first positive phase occurs approximately at a time when a so-called shallow wave will arrive, i.e., $t' = r'$. It is also noteworthy that the maxima, minima, and zeros of corresponding phases of $G_0(r', 0, t')$, $G_1(r', 0, t')$, and $G_1(r', 0, t')$ all have about the same values of t' wherever computed. (This fact seems to hold also for $G_1(r', 0, t')$ where $0.1 \leqslant \tau' \leqslant 1.1$.)

In order to evaluate $G_{z'E,\tau'}$, it is necessary to know the period of the gas globe. The first-period τ can be obtained approximately from the following expression[3] for a spherical globe in an incompressible, inviscid liquid bounded by a

[3] "On the best location of a mine near the sea bed," NDRC, AMP Report 37.1R, AMG-NYU No. 49, May (1944) and Appendix on "Bubble periods" February (1945). Also "On the solution of the differential equations arising from a particular energy function," Mathematical Tables Project, March and October (1944).

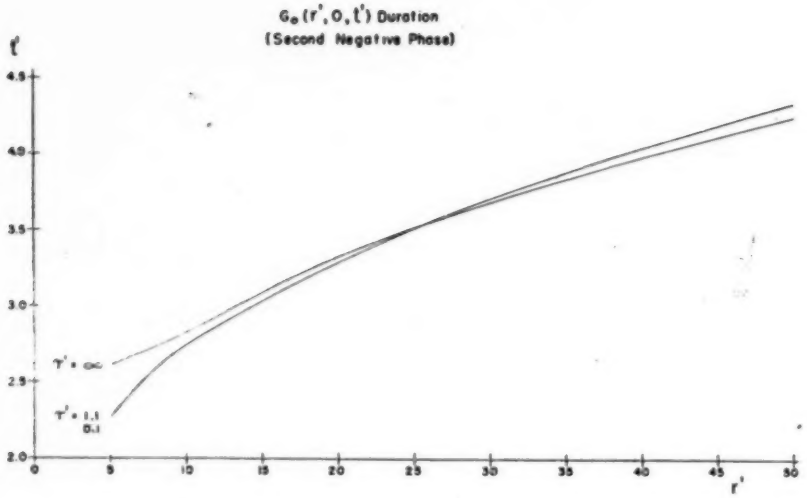


FIG. 3d.

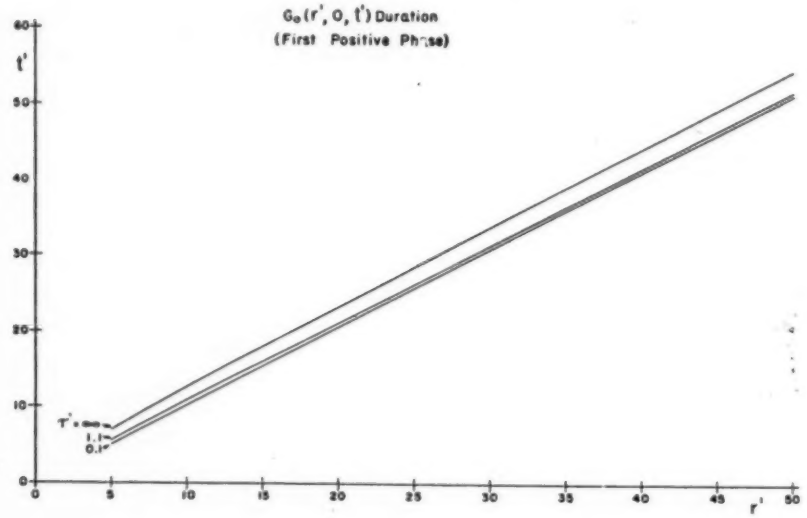


FIG. 3e

rigid sea bed parallel to the free surface, and subject to atmospheric pressure (h_a, in units of length):

$$\tau = 1.485L\left(\frac{3}{2g(h+h_a)}\right)^{\frac{1}{2}}\left(1-\frac{0.189LF(x)}{h}\right), \quad (6)$$

where $F(x)$ is a known function of x and

$$x \equiv 1 - 2z_E',$$

also

$$L \equiv \left(\frac{3E_R}{4\pi\rho g(h+h_a)}\right)^{\frac{1}{3}},$$

where E_R is the amount of residual energy in the gas and the liquid after the shock wave has passed. If E_R is assumed to be 50 percent of the adiabatic, constant-volume heat of explosion, which has been computed[4] to be 1060 cal./g for TNT, then L can be expressed as follows in terms of the charge weight, W, where W is given in pounds and h in feet.

$$L = 14.0\left(\frac{W}{h+33}\right)^{\frac{1}{3}}.$$

Some values of L of practical interest are given in Table I. The expression (6) for the first period of the globe is invalid when the distance of the

[4] J. G. Kirkwood and S. R. Brinkley, Jr., "A new theory of shock wave propagation with an application to the underwater explosion wave produced by TNT," Division 2 Interim Report 24 on "Underwater explosives and explosions," (1944). Confidential.

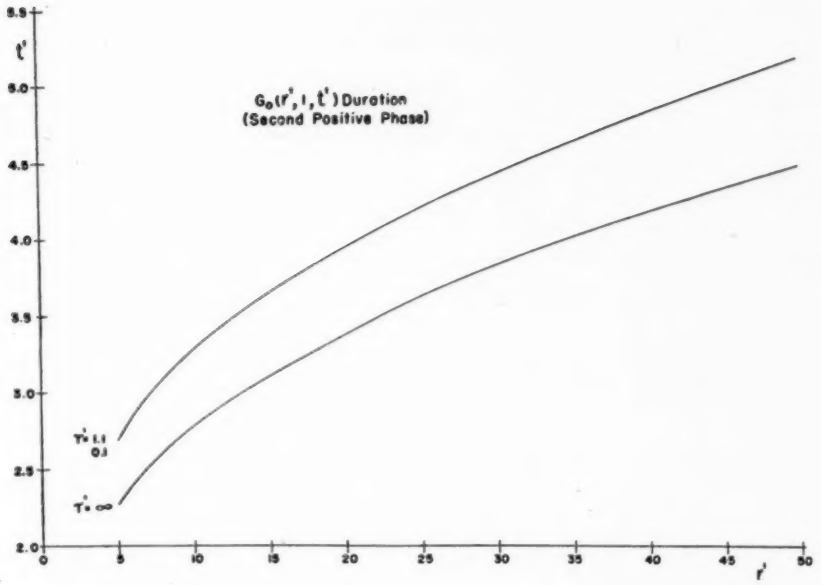


FIG 3f.

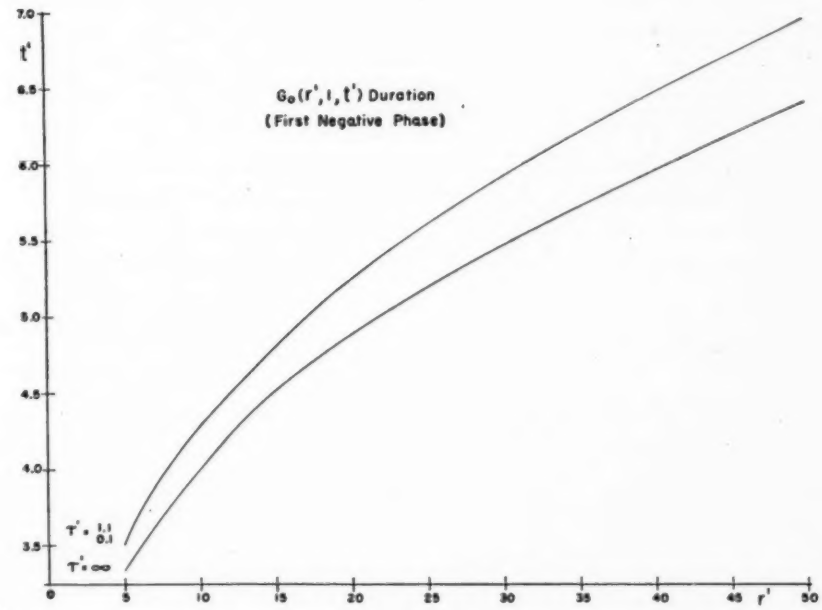


FIG. 3g.

center of the globe form a bounding surface is less than $1.2L$. Nevertheless, in the absence of any other information the value of the first period for $z_E = 1.2L$ will be used heuristically for $z_E \leqslant 1.2L$. On this basis τ and τ' have been computed for various charge weights of TNT over a range of depths; these values are given in Table II. The function $G_{z_E', \tau'}(r', 0, t')$ can be readily obtained from the master function $G_{z'_E}(r', 0, t')$; its characteristic properties for different values of r' are represented for $\tau' = 0.1$ and 1.1 on the various figures (Figs. 2a, 2b, 2c, 2d; 3a, 3b, 3c, 3d; 4).

One additional quantity is required for the determination of the pressure $p_{z_E}(r', z', t')$, namely, the average volume of the gas globe. Thus the time average of the volume for the first period is given by

$$\bar{V} = \tfrac{4}{3}\pi[a(t)]^3,$$

where $a(t)$ is the radius of the bubble at the time t. The dimensionless quantity a/L has been

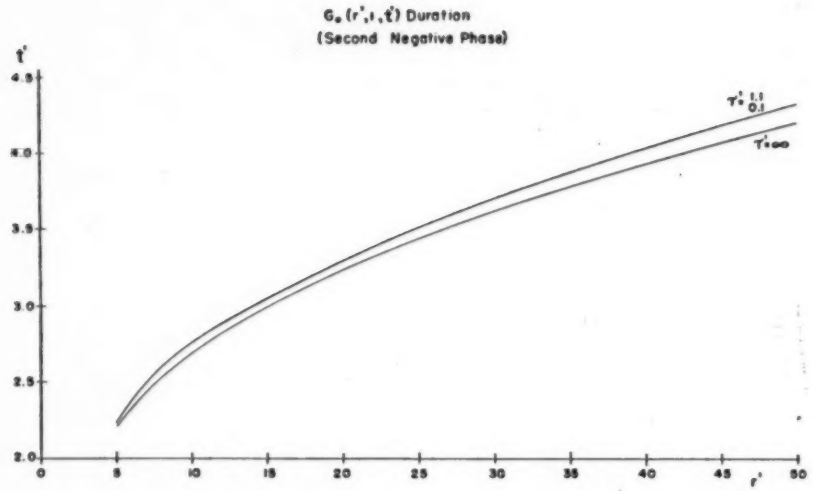


FIG. 3h.

calculated[3] for a globe situated in an infinite liquid with no free surface at a distance of $1.2L$ from a rigid wall. Using these values we find the average of $[a(t)]^3$ over the first period to be 0.4819. Thus we obtain

$$\bar{V} = \tfrac{1}{3}\pi(0.4819)L^3.$$

If the explosive is on the bottom, its gas globe is about half of an oblate spheroid. Small-scale experiments[5] indicate that its size is slightly less than that of the gas globe at a distance of $1.2L$ from the wall. In view of the lack of sufficient quantitative information, however, the actual volume $\bar{V}_B$ on the bottom will be assumed here to be half of $\bar{V}$. (In case of venting at the surface a better approximation is the volume of a cylinder of a height equal to the depth of the water and of a radius slightly less than the maximum spherical globe in an unbounded liquid.)

Thus, we obtain finally for the pressure on the bottom due to the first pulsation from an explosion also on the bottom (see expression (5′))

$$p_0(r', 0, t') = (\rho g/3h^2)(0.4819)L^3 G_{0,r'}(r', 0, t').$$

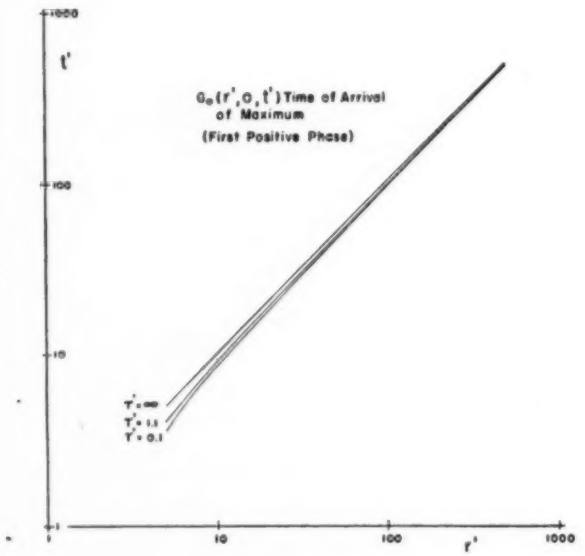


FIG. 4.

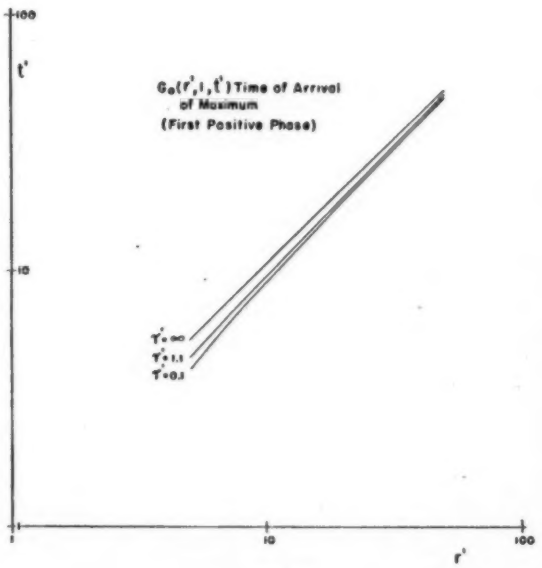


FIG. 4a.

[5] D. C. Campbell, "Motions of a pulsating gas globe under water," David Taylor Model Basin Report 512, May (1943). Confidential.

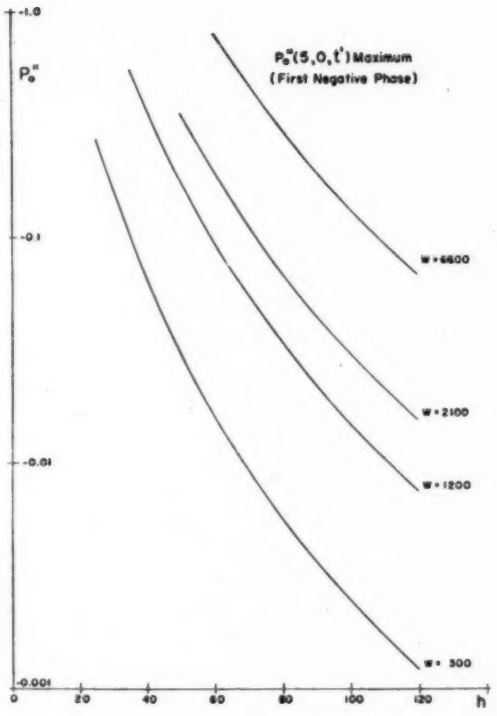


FIG. 5a.

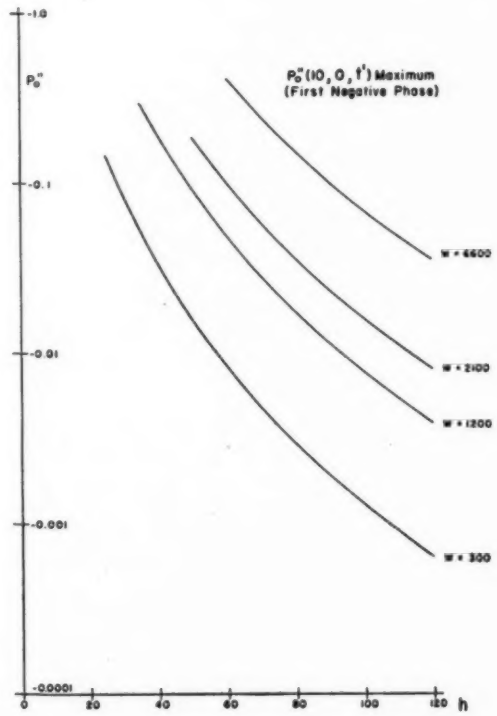


FIG. 5b.

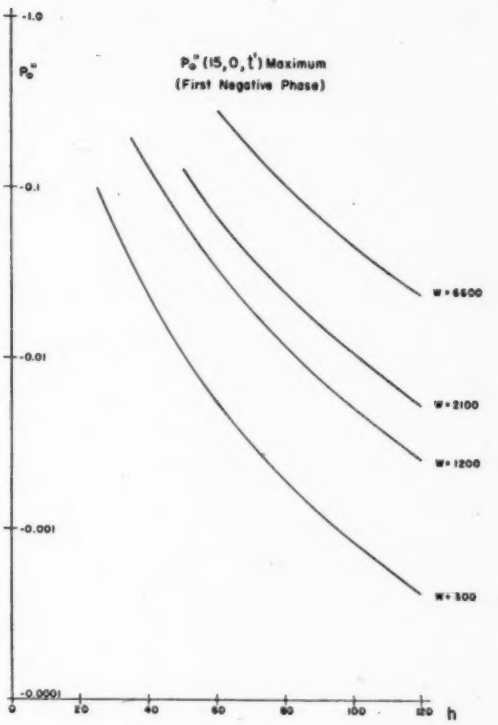


FIG. 5c.

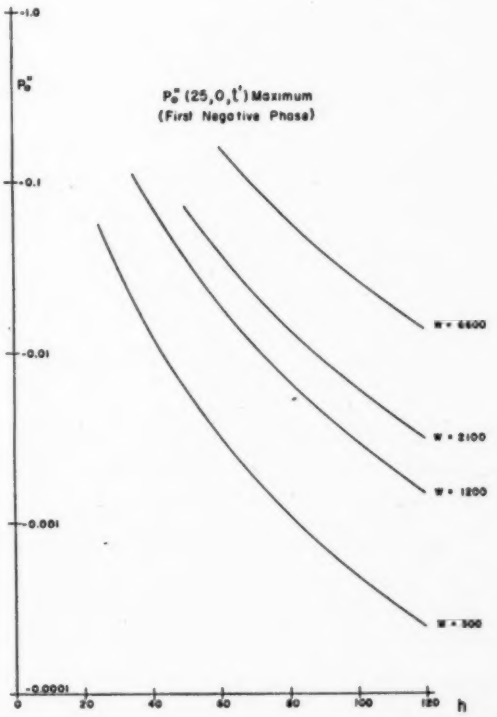


FIG. 5d.

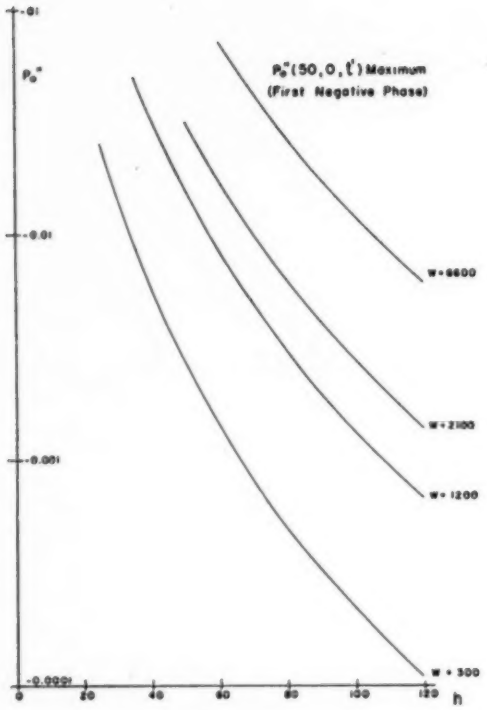


FIG. 5e.

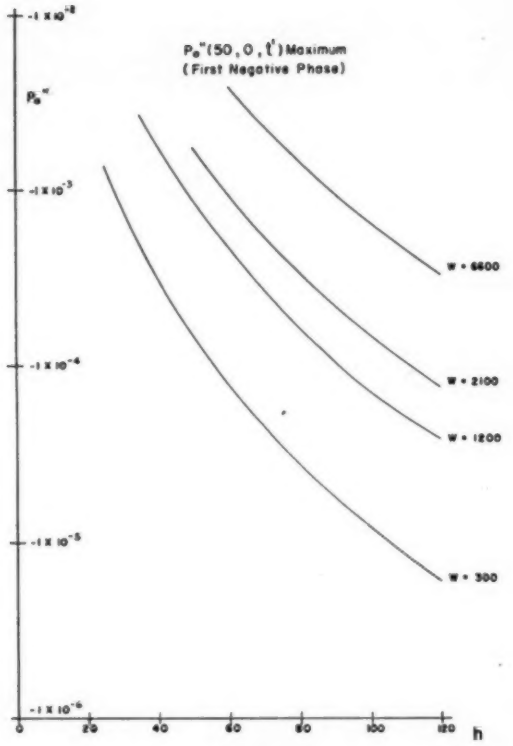


FIG. 5f.

Expressing the pressure as $p_0''(r', 0, t')$ in feet, we have for TNT

$$p_0''(r', 0, t') = 441 \frac{W \cdot G_{0,r'}(r', 0, t')}{h^2(h+33)}.$$

The maximum pressure $p_0''(r', 0, t')$ for the first negative phase is represented as a function of the depth h for various values of the charge weight W in Figs. 5a, b, c, d, e, f, corresponding to values of r' equal to 5, 10, 15, 25, 50, 500, respectively.

A comparison of these calculated values with experimental ones shows satisfactory agreement in consideration of the various approximations that have been made. It will be reported in detail elsewhere.

Calculation of the Temperature Development in a Contact Heated in the Contact Surface, and Application to the Problem of the Temperature Rise in a Sliding Contact

RAGNAR HOLM
Consultant at Stackpole Carbon Company, St. Marys, Pennsylvania
(Received August 15, 1947)

I. A circular area on a semi-infinite body is assumed to be heated at a constant rate from the time $t=0$ on. The development of the temperature in this area and in its surroundings is calculated with certain simplifications.

II. In a sliding contact the heat generated by friction is limited at every point to the time of contact. The temperature reached by the contact surface is calculated by aid of Part I and with assumptions about the size of the contact surface, which are based on earlier investigations of the author.

Observations on the temperatures in bimetallic sliding contacts, indicating the contact temperature rise by means of a thermoelectric current, are compared with calculations partly according to the deduced theory partly according to an earlier published theory by Jaeger. A fairly good agreement is found with one exception.

THE development of the temperature in a constriction region heated by the electric current has already been treated.[1] In that case heat is generated in all points of the constriction region. The present problems are different insofar as the heat is generated only in the contact surface from which it dissipates into other parts of the contact members. A relatively simple solution is found by using similar simplifications as in reference 1, constituting the comparison with a problem with spherical symmetry. The cooling off after heating is finished is also expressed.

The heating problem, particularly the heating by friction, has been treated earlier by H. Blok,[2] J. C. Jaeger,[3] and E. Madelung[4] without the kind of simplifications used by the present author. The mathematical methods were therefore relatively heavy. Jaeger is the only one who has carried out the calculations into considerable numerical detail. We shall make uses of his results.

In considering the problem of the temperature in a sliding contact, we encounter the question of the size and character of the contact surface. This question turns out to be very difficult. Some of the results mentioned in E.C.H.[1] form a basis for the answer, but only a coarse approach to reality can be achieved. It is, therefore, of special interest to compare the theoretical results of Jaeger and of the author with recent measurements by Else Holm.[5] This is done in Table II. A fairly good agreement is noted. The agreement with experimental results by Bowden and Ridler[6] is poorer, possibly because some of the physical constants of the materials are not known with sufficient accuracy.

I. HEATING PROCESSES IN STATIONARY CONTACTS

a. The Development of the Temperature in the Constriction Region of a Stationary Contact Heated in the Contact Surface

We assume a semi-infinite body limited by a plane surface containing the circular area with the radius a in which, from the time $t=0$ on, heat is produced at the constant rate K and

[1] See *Electric Contacts* by Ragnar Holm, (published by H. Geber, Stockholm, 1946). This book is denoted E.C.H. in the following. We here refer to §11. A short definition of the "current constriction" may be given as follows: We distinguish between the apparent and the real contact surface. The real contact surface usually is very small and the lines of flow of the thermal and the electric current starting in or passing through this narrow area are bent together, causing an increased resistance. The region where the streamlines converge or diverge is called the constriction region. The generated surplus resistance is the constriction resistance; cf. E.C.H. §1 and §3.

[2] H. Blok, "General discussion on lubrication and lubricants," Inst. Mech. Eng. London **2**, 222 (1937).

[3] J. C. Jaeger, J. Proc. Roy. Soc. N.S.W. **56**, 203 (1942).

[4] E. Madelung, Naturwiss. **30**, 223 (1942).

[5] Else Holm and Ragnar Holm, Festskrift tillägnad Arvid Hedvall, Chalmers tekn. högskola, Göteborg, Sweden, p. 271, 1948.

[6] F. P. Bowden and K. R. W. Ridler, Proc. Roy. Soc. London **A154**, 640 (1936).

uniformly distributed over the area. No heat may be lost from the plane to the surroundings.

In order to simplify the calculations we imagine the heated, flat surface to be replaced by a hemisphere concentric to it, with the radius b, cf. Fig. 1. This hemisphere may be regarded as an isotherm. Then all isotherms are concentric hemispheres. We shall later discuss the relation between a and b.

The equation defining the development of the temperature is readily found to be

$$\frac{\partial^2\vartheta}{\partial r^2}+\frac{2}{r}\frac{\partial\vartheta}{\partial r}=\frac{c}{\lambda}\frac{\partial\vartheta}{\partial t}, \tag{1}$$

where r is the distance from the center of the b-hemisphere, λ the heat conductivity, and c the specific heat per unit volume of the contact material.

The following boundary conditions are expressions for the assumptions: at $t=0$, $\vartheta=0$ for every r, and at $t>0$ and $r=b$, heat is generated at the rate K uniformly distributed over the b-hemisphere.

Hence,

$$K=-2\pi b^2\lambda\left(\frac{\partial\vartheta}{\partial r}\right)_{r=b}. \tag{2}$$

The substitutions

$$y=\vartheta r-\frac{K}{2\pi\lambda} \quad \text{and} \quad r=b(1+x)$$

transform Eq. (1) into

$$\frac{\lambda}{cb^2}\frac{\partial^2 y}{\partial x^2}=\frac{\partial y}{\partial t}, \tag{3}$$

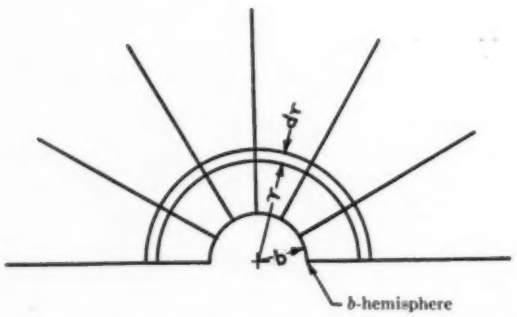


FIG. 1. Streamlines directed radially from the b-hemisphere.

with the conditions that

$$\text{at } t=0,\ y=-K/2\pi\lambda \text{ for every } x,$$

and

$$\text{at } t>0 \text{ and } x=0,\ (\partial y/\partial x)_{x=0}=y(0,t).$$

The solution[7] is

$$-y(x,t)=\frac{K}{2\pi\lambda}\left[\Phi\left(\frac{x}{2(z)^{\frac{1}{2}}}\right)+e^{x+z}\left\{1-\Phi\left(\frac{x}{2(z)^{\frac{1}{2}}}+(z)^{\frac{1}{2}}\right)\right\}\right]=Y(x,z), \tag{4}$$

where

$$\Phi(x)=\frac{2}{(\pi)^{\frac{1}{2}}}\int_0^x e^{-\beta^2}d\beta \tag{5}$$

and

$$z=\frac{\lambda}{cb^2}t. \tag{6}$$

Hence

$$\vartheta(r,z)=\frac{K}{2\pi\lambda r}-\frac{1}{r}Y(x,z). \tag{7}$$

Putting $x=0$, i.e., $r=b$, we find

$$\vartheta(b,z)=\frac{K}{2\pi\lambda b}[1-e^z\{1-\Phi((z)^{\frac{1}{2}})\}]=\Theta(z), \tag{8}$$

designating by $\Theta(z)$ the supertemperature in the contact surface at the time t.

Considering that in the stationary state the temperature may be written const./r, we easily find that according to Eqs. (2) the permanent supertemperature $\Theta(\infty)$ in the b-hemisphere is

$$\Theta(\infty)=\frac{K}{2\pi\lambda b}. \tag{9}$$

Substituting this in Eq. (8) we obtain

$$\frac{\Theta(z)}{\Theta(\infty)}=1-e^z\{1-\Phi((z)^{\frac{1}{2}})\}=B(z). \tag{10}$$

Equation (10) defines the function $B(z)$. Approximation for $z>4$:

$$B(z)=1-\frac{1}{(\pi z)^{\frac{1}{2}}}\left(1-\frac{1}{2z}\right).$$

[7] See Riemann-Weber *Die Partiellen Differentialgleichungen der Mathematischen Physik* (Braunschweig, 1912), Volume II, especially p. 97 and p. 98.

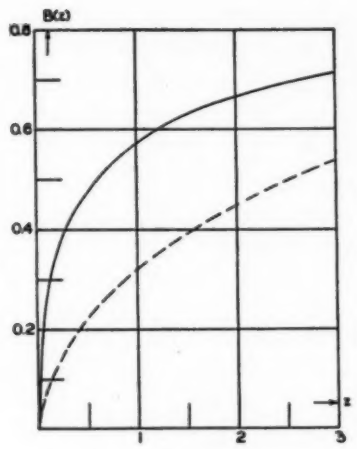


FIG. 2.

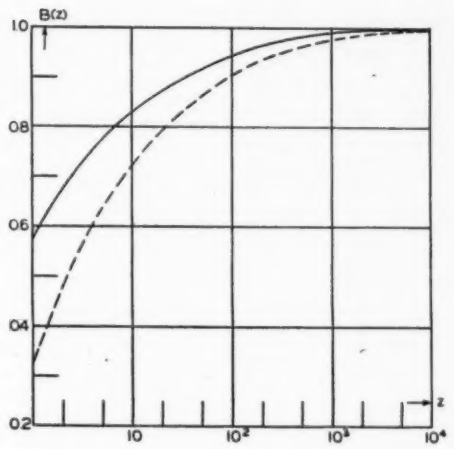


FIG. 3.

FIGS. 2 AND 3. The full drawn curves illustrate the function $B(z)$ of Eq. (10). The dashed curves refer to a contact heated by the electric current through the contact and are taken from the book E.C.H., §11.

The function $B(z)$ is illustrated by the full-drawn curves in Figs. (2) and (3) differing in that the second figure has logarithmic abscissae. The dashed curves present the case of a contact heated by the electric current.[8] A striking feature of the new curve is its beginning with a vertical tangent.

b. The Constriction Region Cooling Off

We assume that the heating described above has reached the stationary state, and that now at the time $t=0$ heating ceases. Then Eq. (1) is still valid but with the following conditions:

$$\text{at } t=0,\ \vartheta = K/2\pi\lambda r$$

and

$$\text{at } t>0,\ (\partial\vartheta/\partial r)_{r=0}=0.$$

The substitutions

$$y=\vartheta r \quad \text{and} \quad r=b(1+x)$$

transform Eq. (1) into

$$\frac{\lambda}{cb^2}\frac{\partial^2 y}{\partial x^2}=\frac{\partial y}{\partial t},$$

with the conditions that

$$\text{at } t=0,\ y=K/2\pi\lambda$$

and

$$\text{at } t>0,\ (\partial y/\partial x)_{x=0}=y(0,t).$$

The general solution is

$$y(x,t)=Y(x,z)$$

or, denoting the decreasing supertemperature by $\vartheta_c(r,z)$ or $\Theta_c(z)$, respectively, where the index c refers to the cooling process,

$$\vartheta_c(r,z)=\frac{1}{r}Y(r,z)=\frac{K}{2\pi\lambda r}-\vartheta(r,z), \qquad (11)$$

and in the special case $r=b$, i.e., for the contact surface

$$\Theta_c(z)=\Theta(\infty)e^z\{1-\Phi((z)^{\frac{1}{2}})\}. \qquad (12)$$

II. THE TEMPERATURE RISE IN A CONTACT SPOT[9] OF A SLIDING CONTACT BECAUSE OF THE FRICTION

a. Introduction

We regard a circular contact area moving with the constant speed v along the surface of the base contact member. The friction produces heat at the constant rate K in it, uniformly distributed over the area. We want to find the generated temperature rise above the temperature of distant parts of the contact member which are considered as remaining at room temperature. The rise will not be the same in all points of the contact area, and we shall first try to compute the maximum temperature rise. A computation of an average temperature is also important.

[8] See E.C.H., case A in Figs. (11.14) and (11.15).

[9] The contact surface between practically plane surfaces is usually subdivided into small spots, which in E.C.H. are called a-spots if they represent a metallic contact.

We proceed in the following way. First we make the supposition that the heat is produced in a stationary contact spot. After a long time ($t=\infty$) the supertemperature attains a constant value, corresponding to $\Theta(\infty)$ in Eq. (9).

Then we take into account that the heat production in any contact point of the base sliding member is transient, and that therefore only a fraction B of the supertemperature $\Theta(\infty)$ will be reached. We shall find that we may use Eq. (10) for the calculation of B if we make a suitable choice of the quantity b for Eq. (6).

We remember that Eqs. (9) and (10) are deduced for the case that the heated surface is hemispherical. We must find the conversion to the case of the flat heated spot.

Eq. (10) still gives a fairly good approximation of B. This may be proved by a simple application of the discussion in E.C.H., §11*C*, noticing that the quantity b is defined by Eq. (19).

The substitute for Eq. (9) requires more discussion. We consider the permanent temperature to be the thermomotive force which forces the heat current across the thermal constriction resistance W connected with the contact spot.

If $\Theta_m(\infty)$ means the maximum temperature in the contact spot with the radius a, i.e., the temperature in its center, $\Theta_m(\infty)$ may be compared with the electric potential (above the potential in infinitely distant points) of the center of a circular disk with the same radius and the entire charge $K/2\pi\lambda$ uniformly distributed. The number K/λ of lines of electric force which start from one side of the disk multiplied by the conductivity λ then represents the thermal current K in our problem. The electric potential is known to be $K/\pi a\lambda$. Hence[10]

$$\Theta_m(\infty)=\frac{K}{\pi a\lambda} \quad \text{and} \quad W_m=\frac{1}{\pi a\lambda}. \tag{13}$$

We shall meet the question about the temperature which is measured when the sliding contact, with members of different metals, acts as a thermocouple. Then, not $\Theta_m(\infty)$ but a kind of average temperature is responsible for the thermocurrent. It probably will not be the arithmetical average. Another possible average temperature $\bar{\Theta}(\infty)$ may be defined as that temperature in the a-spot which would provide the heat current K if the spot were isothermal. In this case the thermal resistance is known[11] to be $1/4a\lambda$ and thus

$$\bar{\Theta}(\infty)=\frac{K}{4a\lambda}. \tag{14}$$

We have

$$\frac{\Theta_m(\infty)}{\bar{\Theta}(\infty)}=\frac{4}{\pi}=1.27, \tag{15}$$

which is just the order Jaeger, in another way, calculates for the ratio between the maximum and arithmetical average temperature in a rectangular contact spot.

b. Calculation of K and the Radius a of a Flat Contact Spot

Let P be the entire contact load and n the number of a-spots. Thus, the load on one a-spot is P/n. Designating the coefficient of friction by μ, we obtain for the single a-spot

$$K=\mu\frac{Pv}{n}. \tag{16}$$

With the units used below, P in g, v in cm/sec., and energy in joules, we get

$$K=10^{-4}\mu\frac{Pv}{n}. \tag{17}$$

The calculation of the radius a must be based on the experimental experience that the average pressure in a contact usually is of the order of the hardness H of the softer member.[12] In crossed-rod contacts H is often reached, in plate contacts the pressure usually is of the order $H/2$. The contact between a small slider and a base plate constitutes an intermediate case, and we assume the average pressure to be $0.7H$. Thus, by definition of pressure,

$$P=n\pi a^2 0.7H$$

or

$$a=\left(\frac{P}{0.7n\pi H}\right)^{\frac{1}{2}}. \tag{18}$$

c. Calculation of the Function z

Equation (6) contains the quantity b which represents the linear dimension of the contact

[10] Equation (13) corresponds to Jaeger's Eq. (31).

[11] See E.C.H., §5K.

[12] See E.C.H., §14 and the beginning of §40B.

spot and thus will be proportional to a. The principle for determining the relation between a and b must be that the heat generated at the rate K may cause the same average stationary temperature in the real plane spot with the radius a and in the hypothetical hemisphere with the radius b. Hence, according to Eq. (9),

$$\frac{K}{2\pi\lambda b}=\frac{K}{4a\lambda}$$

or

$$b=a\frac{2}{\pi}. \tag{19}$$

Equation (6) also contains the time t during which the contact spot is considered to be heated. In the assumed case of a circular heat source moving with the speed v on the surface of the base contact member, t will be the average time during which a point of this member has contact with the slider. The average length of the contact spot in the v direction is $a\ \pi/2$. Hence,

$$t=\frac{\pi}{2}\frac{a}{v}. \tag{20}$$

According to Eqs. (19) and (20) we obtain

$$z=\frac{\pi^3}{8}\frac{\lambda}{cav}. \tag{21}$$

Using z according to Eq. (21), $B(z)$ is determined by Eq. (10).

d. Course of the Numerical Calculations

K and a are obtained by means of Eqs. (17) and (18). Then Eq. (13) or Eq. (14), respectively, determine $\Theta_m(\infty)$ or $\bar{\Theta}(\infty)$. The quantity $B(z)$ is obtained from Eq. (10) with z according to Eq. (21). Finally, we obtain

$$\bar{\Theta}(z)=B(z)\bar{\Theta}(\infty) \quad \text{or} \quad \Theta_m(z)=B(z)\Theta_m(\infty). \tag{22}$$

e. Application to Actual Contacts

The simplest case for the application of the theory is realized if one contact member, a relatively small slider, glides on a plane base member generating a single circular contact surface with a constant position on the slider, and if the slider conducts heat only to a negligible amount.

Experiments I

Bowden and Ridler[6] have made experiments which seem to correspond to the described case. They used a small pin slider gliding with its end on a flat plate. Bimetallic contacts were used and the thermoelectric voltages measured. They calibrated their thermocouple by assuming that the highest temperature, observed to be an upper limit, was just the melting point of the one contact member.

Melting in the contact begins when the temperature in its center attains the melting point, and the described method of calibration will nearly correspond to adapting the device for indicating the maximum temperature in the contact. We, therefore, use Eq. (13) for calculations concerning these experiments I. When applying Jaeger's theory we use his Eq. (33), giving the maximum temperature in a square contact spot.

Experiments II

Else Holm[5] has made similar measurements. The sensitivity of the thermocouple constituted by the sliding contact was calibrated by means of other couples consisting of the same materials with one juncture dipping in melting snow and with the other in water at room temperature or in initially dry Na_2SO_4 just combining with a sufficient quantity of water. The recorded temperature is an average for the contact surface and is to be calculated according to Eq. (14), or as the arithmetical average when using Jaeger's theory.

The slider was not small enough to be neglected in the thermal balance, but its ability to conduct heat from the contact was limited. A good approximation will be to assume that the slider functions as if its thermal conductance were half of that of a semi-infinite contact member of

TABLE I. Physical constants.

Metal	λ cm^{-1} deg.$^{-1}$	c j cm^{-3} deg.$^{-1}$	H 10^6 g cm^{-2}
Ni	0.6	3.9	16.6
Fe*	0.45	3.6	—
Brass	0.9	3.2	8.3
Sn	0.64	1.6	0.5
Pb*	—	—	1.0

* Assumed values for the experiments I.

TABLE II.

Contact	v cm/sec.	$\bar{\Theta}(\infty)$	z	$B(z)$	$\bar{\Theta}(z)$ calcul. accord. to Holm	$\bar{\Theta}(z)$ calcul. accord. to Jaeger*	$\bar{\Theta}(z)$ observed
Experiments II (Holm), $\mu=0.7$							
Ni-brass $P=450$ g $n=3$ $a=2.86\times10^{-3}$ cm	9	7.9	23	0.88	7.0	10.5 Eq. (40)	8.8
	15	13.2	14	0.85	11		11.0
	35	30.7	6	0.79	24		20.5
	80	70	2.6	0.70	49	67 Eq. (44)	38.0
Ni–Sn $P=450$ g $n=1$ $a=20.2\times10^{-3}$ cm	23	9.8	1.3	0.61	6.0		5.0
	35	15	0.84	0.55	8.3		6.0
	80	34	0.37	0.42	14.2		10.0
Ni–Sn $P=1130$ g $n=1$ $a=32\times10^{-3}$ cm	13	8.7	1.5	0.63	5.5		4
	30	20	0.65	0.51	10		8.0
	80	54	0.24	0.35	19	24 Eq. (42)	13.6
Experiments I (Bowden and Ridler), $\mu=1$							
		$\Theta_m(\infty)$				$\Theta_m(z)$	$\Theta_m(z)$ obs.
Pb–Fe $P=57$ g $n=1$ $a=4.25\times10^{-3}$ cm	500	475	0.23	0.35	166	185 Eq. (33)	209
Pb–Fe $P=102$ g $n=1$ $a=5.7\times10^{-3}$ cm	160	202	0.53	0.48	97	120 Eq. (33)	200
	600	762	0.14	0.28	214	232	300

* Jaeger deduces different formulae for maximum and average temperatures and uses different approximations for different sliding speeds.

the same metal. This will be taken into account in two different ways.

When using Eq. (14), which is deduced for heat conduction through one contact member (the base member), we add half the conductivity λ of the slider to the conductivity of the base member. When using Jaeger's Eqs. (42) and (44) we just set the conductivity of the metal of the slider half as large as its normal value.

The investigators have not observed the coefficient of friction simultaneously with the measurement of the temperature. We, therefore, have to make plausible assumptions about μ. Since the wear in the contacts was considerable, the sliding surfaces, on an average, probably were nearly clean metallic surfaces. In such surfaces the friction force per unit area[13] usually is of the order $H/2$, where H is the hardness. In a contact area S the entire friction is $(H/2)S$. Thus,

$$\mu = HS/2P.$$

Since we suppose $p=P/S=0.7H$, we obtain

$$\mu \approx 0.7.$$

We use this value for calculations on experiments II. For the experiments I we just set $\mu=1$, cf. below.

As mentioned above, the slider in the experiments I may have had one single contact spot with a constant position on the slider. But in some of the contacts of the experiments II the contact surface was more complicated. It may have been subdivided into spots and their position on the slider may have varied. We have chosen the number n of contact spots for Table II so as to give a good agreement with the observations. The number n is also plausible when compared with other experiences. The variation of the position of the spots is already taken into account by using Eq. (20) for t. Because a corresponding step is not taken when using Jaeger's equations, we expect them to give a little too high temperatures, as seen in Table II.

In the use of Jaeger's formulae, which are valid for a square contact area with the side $2a$, we assume the heat per unit area to be the same as in the circular area, i.e., equal to $(K/\pi a^2)$. This corresponds to a greater total heating rate than K, but a closer agreement between calculations with both forms of the contact surface is achieved in this way.

Table II summarizes the measurements and calculations. The material constants used for these calculations are recorded in Table I.

Table II shows a fairly good agreement between the results of the experiments II and the calculations. A typical example of the experiments I is represented in the lower part of Table II. Here an agreement between calculations and observations could be obtained by using rather extreme values of the material constants, putting $\mu=1$, and by replacing the factor 0.7 in Eq. (21) by 1. It would be valuable to get more experimental information on this point.

[13] Cf. E.C.H., §38.

Measurement of Carbon Black Particles by the Electron Microscope and Low Temperature Nitrogen Adsorption Isotherms

R. B. ANDERSON* AND P. H. EMMETT**
Chemical Engineering Department, Johns Hopkins University, Baltimore, Maryland
(Received September 6, 1947)

Electron micrographs and nitrogen adsorption isotherms have been compared on six commercial carbon blacks as methods for obtaining particle size and surface area values of finely divided materials. Good agreement between the two procedures was obtained for four of the backs: Grade 6, P 33, acetylene black, and lampblack. For two blacks known to be porous, Mogul and after-treated lampblack, the areas, as measured by the nitrogen isotherms, were 5 to 7 times as large as one would calculate from the electron microscope results.

DURING the past ten years the Columbian Carbon Company has made an extensive study of the size distribution of the particles of a variety of carbon blacks by means of the electron microscope.[1,2] On calculating apparent surface areas from the size distribution curves and comparing them with areas that had been obtained by others by adsorption methods, they reported that "adsorption values are unquestionably influenced by the type of surface activity present on the carbons,"[2] and that "adsorption methods (both for iodine and nitrogen) appear unreliable as a guide to surface area for the complete carbon family."

In disagreement with this conclusion, De Witt and one of the authors[3] have called attention to the almost perfect agreement that is obtained between the average particle size of Micronex, as measured by the electron microscope and as calculated from the surface area determination by the low temperature nitrogen adsorption method, that has proved to be so widely applicable and generally accepted as a means of measuring the surface areas of a variety of finely divided or porous adsorbents.[4–7] Furthermore, the present authors[8] have established that typical carbon blacks such as Grade 6 or Carbolac 1 have areas by the nitrogen adsorption method that are quite independent of the amount of surface complex and hence of the "surface activity" of the blacks. Accordingly, there seems good reason to believe that areas estimated by the electron microscope and by nitrogen adsorption should agree for all non-porous carbon blacks and that any disagreement can hardly be attributed to the influence of surface activity on the nitrogen adsorption methods.

To obtain a better and more convincing comparison of the two methods, the present work was undertaken. All previously reported comparisons had been made on samples of commercial blacks obtained in different locations and frequently at widely separated times. It seemed desirable to make surface area measurements by the nitrogen adsorption method and particle size distributions by the electron microscope method on samples of several types of commercial carbon blacks under carefully controlled sampling conditions. Accordingly, identical samples of six different carbon blacks were studied by the two methods. The results seem to show that there is good agreement between the two methods on all non-porous carbon blacks. Only on samples that have been purposely made porous by air or steam activation processes are divergent results obtained. This, however, is to be expected since the electron microscope method measures average particle size and particle distribution whereas the nitrogen adsorption method gives the total sur-

* Present address: Central Experiment Station, U. S. Bureau of Mines, Forbes Street, Pittsburgh, Pennsylvania.
** Present address: Mellon Institute, Fifth Avenue, Pittsburgh, Pennsylvania.

[1] *Columbian Colloidal Carbons*, Volume II, 1940.
[2] *Columbian Colloidal Carbons*, Volume III, 1942.
[3] Emmett and DeWitt, Ind. Eng. Chem., Anal. Ed. **13**, 28 (1941).
[4] Emmett, J. Am. Chem. Soc. **68**, 1784 (1946).
[5] Wooten and Brown, J. Am. Chem. Soc. **65**, 113 (1943).
[6] Beebe, Beckwith, and Honig, J. Am. Chem. Soc. **67**, 1554 (1945).
[7] Ries, Van Nordstrand, Johnson, and Bauermeister, J. Am. Chem. Soc. **67**, 1242 (1945).
[8] Emmett and Anderson. J. Am. Chem. Soc. **67**, 1492 (1945).

face area accurately, but yields correct values for average particle size only when the adsorbent is non-porous.

EXPERIMENTAL

The present work was restricted to four non-porous and two porous commercial carbon blacks. All samples were furnished by the Godfrey L. Cabot Company. The non-porous blacks were P33, acetylene black, lampblack, and Grade 6 rubber black. The porous samples were Mogul and lampblack LBTT (after-treated lampblack T). Mogul is a standard long ink black prepared commercially by the after treatment in air at 800°F of a short ink black of the Elf type.[9]

Values for the surface areas were obtained by measuring the adsorption of nitrogen at −195°C at a number of points over the relative pressure range 0.05 to 0.5, and plotting the data according to the standard Brunauer-Emmett-Teller[10] equation

$$\frac{x}{V(1-x)}=\frac{1}{V_mC}+\frac{(C-1)x}{CV_m}, \qquad (1)$$

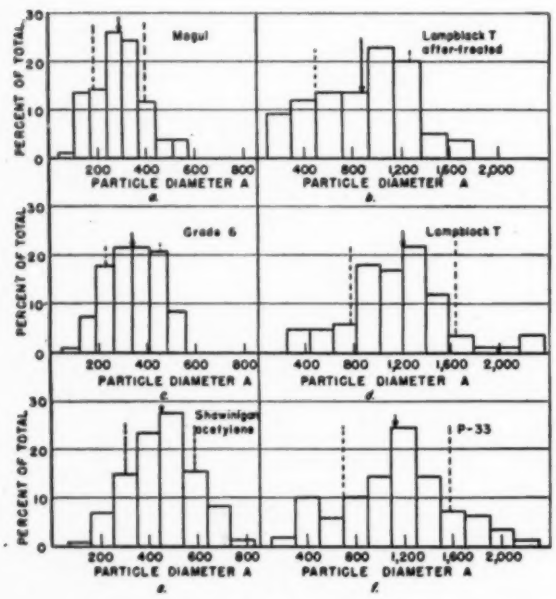


FIG. 1. Histograms of the particle size distribution of various carbon blacks from electron micrographs. The arrows indicate the arithmetic means, and the vertical dotted lines are drawn one standard deviation unit on each side of the mean.

where V is the volume of gas adsorbed (as measured at standard temperature and pressure) at the relative pressure x, C is a constant, and V_m is the volume of gas required to form a monolayer on the surface of the adsorbent. By assuming that the cross-sectional area of the adsorbed nitrogen molecule is 16.2A², one can obtain from V_m the area of the sample in square meters by multiplying V_m by 4.38.[11] The details of the adsorption method and the reasons for believing that the method furnishes reliable surface area for finely divided materials have been described[3, 4, 10, 12] elsewhere and need not be repeated here.

The electron microscope photographs were taken by Dr. Hall of the Massachusetts Institute of Technology; he used[13] standard "film" preparations[14] and the procedure that has become customary for such materials. The photographs were actually taken at a magnification of about 9000. The measurements were made from prints enlarged to a total magnification of from 30,000 to 70,000 depending on the black. On the electron microscope prints, most of the particles appeared to be spheres or groups of spheres, though, admittedly, the photographs of acetylene black were not sufficiently detailed to preclude the possible presence of some crystalline platelets of the type recently described by Watson.[15, 16] Only particles with circular cross sections in which more than half of the perimeter was visible were measured to avoid counting clusters as single particles. The magnified images of the particles were measured on the prints to the nearest 0.5 mm.

RESULTS

In Fig. 1 are shown the particle size distribution plots for the six carbon blacks. The mean value of the diameter is indicated by an arrow on each of the plots, the standard deviation is indicated by parallel verticle dotted lines, one on each side of the arrow.

In Table I is presented a summary of data for the electron microscopic counts and for the sur-

[9] Emmett and Cines, J. Phys. and Colloid. Chem. **51**, 1329 (1947).
[10] Brunauer, Emmett, and Teller, J. Am. Chem. Soc. **60**, 309 (1938).
[11] Emmett and Brunauer, J. Am. Chem. Soc. **59**, 1553 (1937).
[12] Emmett, Am. Soc. Testing Materials, "Symposium on new methods for particle size determination in subsieve range," p. 95 (1941).
[13] Hall, private communication.
[14] Watson, J. App. Phys. **18**, 153 (1947).
[15] Watson, J. App. Phys. **17**, 121 (1946).
[16] Watson, Electrochem. Soc. **92**, 4 (October, 1947).

TABLE I.

Sample	From electron micrographs: Number of particles counted	Magnification	d_n A	d_A A	Calculated area m^2/g $\rho=1.85$	Calculated area m^2/g $\rho=2.00$	From nitrogen adsorption: Calculated d_A in A $\rho=1.85$	Calculated d_A in A $\rho=2.00$	Area m^2/g	Roughness factor*
Grade 6	82	51,580	345	410	79.1	73.2				
	105	66,800	340	400	80.6	74.6	295	273	110	1.43
P33	110	50,000	1130	1430	22.7	21.0				
	63	31,900	1220	1460	22.2	20.6	1565	1459	20.7	0.96
Shawinigan acetylene	117	52,100	435	540	60.2	55.6				
	140	51,000	450	500	64.6	59.8	502	465	64.5	1.08
Lampblack *T*	83	54,090	1200	1520	21.3	19.8				
	64	54,090	1010	1280	25.3	23.4	1297	1200	25.5	1.07
	42	34,000	1040	1230	26.4	24.4				
Lampblack LBTT (after-treated lampblack *T*)	51	30,400	810	1040	31.1	28.8				
	139	48,000	880	1150	28.1	26.0	156	144	208	7.30
Mogul	152	73,900	290	360	89.9	83.2	72	67	450	4.20

* The roughness factor is merely the ratio of the surface area as measured by N_2 to the area calculated from d_A of the electron microscope results using a density of 1.925 for the carbon blacks. This average value for density is employed because experimental density values were not available for all of the carbon blacks used. However, to show the influence that a variation on the values for the true density might have on a comparison of the nitrogen adsorption and electron microscope methods, calculations in columns 6 to 9 have been made for density values that may be considered the upper and lower limits usually found for carbon blacks.

face area measurements. Column 2 gives the number of particles included in each of the counts; column 3, the magnification; column 4, the mean diameter of the blacks d_n defined as $(\Sigma nd/n)$; and column 5 gives values for d_A defined as $\Sigma nd^3/\Sigma nd^2$. The surface area values are shown directly in column 10 as square meters per gram. As pointed out above, these areas were obtained in each case by evaluating V_m from Eq. (1), and then multiplying the V_m value by 4.38 and dividing by the weight of the sample to convert the result into square meters per gram.

The surface area values calculated from the electron microscope and the average particle size values calculated from the surface area measurements involve the assumption of some shape factor[16,17] for the particles, and of a value for the density. The surface area calculations from the electron microscope photographs and the size calculations from the adsorption areas have been made on the assumption of spherical particles of a density of 1.85 and also of a density of 2.00 for the carbon blacks. These calculations are shown in columns 6, 7, 8, and 9.

DISCUSSION

Critical discussions by various previous workers[17] have made it clear that to obtain the area of a collection of spherical particles for which a size distribution curve is available, the relation

$$\text{Area} = 6/(\rho d_A) \qquad (2)$$

may be used, where ρ is the true density of the particles and d_A, as defined above, is obtained by dividing the sum of the cubes of the measured diameters of the individual particles by the sum of the squares. Conversely, this same equation may be used for obtaining values of d_A, provided the total surface area (as measured for example by the nitrogen adsorption method) and the particle density are known.

The comparisons between the electron microscopic and the gas adsorption methods may be made in several ways. Column 11 of Table I lists a roughness factor calculated as the ratio between the surface area as measured by nitrogen and the surface area calculated by Eq. (2) from the electron microscope distribution curves using the d_A values from column 5 and the average of the two density values 1.925 for the density of the carbon. The results show clearly that the agreement between the electron microscope and the nitrogen adsorption method is well within the expected experimental error*** for the two

[17] Green, J. Frank. Inst. **204**, 713 (1927).

*** Strictly speaking, only three of the four non-porous blacks yield areas which agree by the two methods within expected experimental errors. The fourth carbon black,

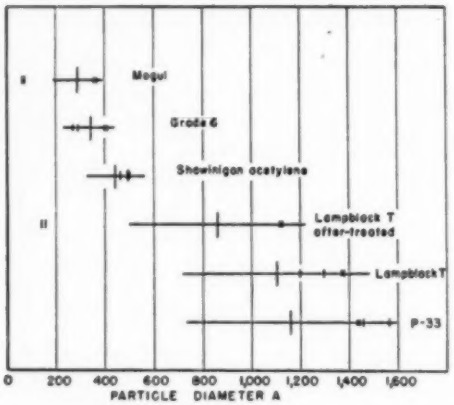


FIG. 2. A plot of the means and standard deviations from electron micrographs compared with the average, d_A, computed from nitrogen adsorption isotherms. The averages d_n and d_A from the micrographs are represented by the long vertical line and X, respectively. The horizontal line extends one standard deviation unit on each side of d_n. The two short vertical lines represent the values of d_A computed from nitrogen surface areas with carbon black particles assumed to be smooth spheres with densities of 1.85 and 2.00.

methods on the four non-porous carbon blacks. The area calculations from the electron micrographs probably involve an experimental error of at least 10 percent. The surface area values from the nitrogen adsorption are reproducible to a few percent, but are uncertain in an absolute sense to perhaps 20 percent because of the uncertainty in the cross-sectional area values chosen for the adsorbed nitrogen molecules. As has been pointed out in previous publications,[11] the area that one calculates for the cross-sectional area of the nitrogen molecule is $13.7A^2$ if the calculation is based on the density of solid nitrogen, and $16.2A^2$ if the density of liquid nitrogen is used in the calculation. Evidence seems to be accumulating[3,18,19] to indicate that the average area occupied by the adsorbed nitrogen molecule is closer to the 16.2 value than to the 13.7 value, but this may well vary somewhat from solid to solid. In view of these possible combined uncertainties, the results for the four non-porous carbon blacks, as determined by the two methods, may be considered to be in good agreement.

Grade 6, has a roughness factor of 1.43 which is just about the limit of combined errors of the two methods. Larger numbers of particles of this black would have to be counted to be sure that this value represents a real roughness of the particles.

[18] Harkins and Jura, J. Am. Chem. Soc. **66**, 1362 (1944).

[19] Harkins and Jura, J. Am. Chem. Soc. **66**, 1366 (1944).

The two porous blacks have been shown elsewhere[9] to have isotherms resembling those obtained for charcoal and characteristic of substances with a considerable fine pore structure. The roughness factors of 7.3 and 5.2 are, accordingly, to be expected.

The results shown in Table I are illustrated in two separate plots. Figure 2 shows by a long vertical line the mean diameter of the particles observed for the different blacks by measurement of the electron micrographs. The horizontal lines are drawn one standard deviation unit on each side of the mean. For a normal distribution about two thirds of the particles lie within the size range covered by this horizontal line. Also, for each black the value for d_A from the electron micrographs is plotted in Fig. 2 by use of the symbol X. Finally, by short vertical lines are indicated the limits to the value of d_A as calculated from the surface area measurements for carbon densities of 1.85 and 2.0. In Fig. 3, the surface areas as measured directly by nitrogen are plotted against those calculated from the d_A values of the electron micrographs using Eq. (2),

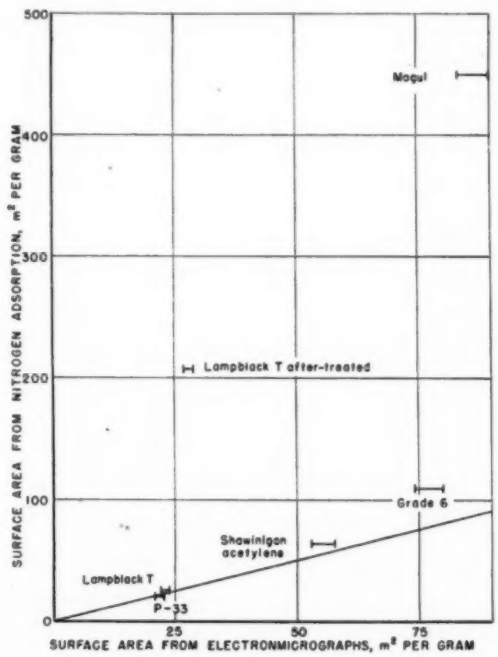


FIG. 3. Comparison of surface areas from nitrogen adsorption to areas computed from electron micrographs assuming the particles to be smooth spheres. The two short vertical lines |—| represent the areas computed from micrographs with the density of the carbon blacks assumed to be 1.85 and 2.00.

TABLE II. Comparison of electron microscope and nitrogen adsorption values* obtained by various workers on carbon blacks.

Reference	Micronex (*M*)** or Grade 6 (*G*)**			P33			Acetylene black			Lampblack *T*			Lampblack LBTT (after-treated lampblack *T*)			Mogul		
	d_n	d_A	Area	d_n	d_A	Area	d_n	d_A	Area	d_n	d_A	Area	d_n	d_A	Area	d_n	d_A	Area
Columbian Carbon Book III†	280*M*			740			430											
Columbian Carbon Co. Weigand	280*M*			740			430											
Columbian Carbon Co. Weigand and Ladd	280*M*			740			430											
B. F. Goodrich Co.	326*G* 286*M* 247*M* 264*M*	430 356 372 372		1333	1475		360	515		1188	1624							
Emmett and DeWitt	362*M*	480	107			22.1			64.5									
Emmett and Cines			110 112									26			218			475
Godfrey L. Cabot, Inc.	286*G*			1171			426			926								
Godfrey L. Cabot, Inc.—Smith, Thornhill, and Bray			114			15.0			64			28						{450 {350
Watson	295*M*-6			1050			430											
Anderson and Emmett (this work)	341*G*	406	110	1170	1440	20.7	445	500	64.5	1110	1380	25.5	860	1130	208	290	360	450

* d_n and d_A are from electron microscope photographs and are, respectively, $(\Sigma nd/n)$ and $(\Sigma nd^3/\Sigma nd^2)$ expressed in angstrom units; areas are in square meters per gram as measured by nitrogen adsorption isotherms.
** *M* and *G* are used in column 2 to indicate, respectively, Micronex and Grade 6 carbon black.
† The papers in this column are cited in references 2; 24; 25; 21; 3; 9; 20; 22; 14 and 16 (both Watson), respectively.

the two short vertical lines indicating the areas calculated from the micrographs with the carbon densities taken as 1.85 and 2.0. The agreement, as pointed out above, is all that could be desired for three of the four non-porous blacks and is fairly good for the fourth; only the two porous blacks give larger areas by the nitrogen method than by the electron microscope method.

The question naturally arises as to how the results obtained by our surface area and our electron microscope measurements compare to those previously published for these blacks by other workers. In Table II, all of the other known results for these blacks are summarized. It is apparent that the electron microscope results obtained in the present work agree well with others that have been reported, except for P33. For this black the Columbian Carbon group[2] report a d_n value of 740A, which is about 40 percent smaller than the average of the values obtained in the present work and those by three other workers.[14, 20, 21] It seems probable that the d_n value should be in the range 1000 to 1300 rather than in the range 700 to 800 for this black. For the nitrogen surface area measurements, the values obtained in the present work agree well with those reported by others,[3, 22] except for the one value of 15 square meters per gram reported by Smith, Thornhill, and Bray[22] for the area of P33. Certainly, areas obtained in the present work and in the paper of Emmett and DeWitt[3] for P33 agree much better with the electron microscope counts than does the 15 square meter value. The cause of the P33 area measurement being low in the work of Smith, Thornhill, and Bray, and of the electron microscope value being too low in the work of the Columbian Carbon, is not known. Possibly variations among the various batches of P33 might be large enough to account for the spread of results. For the most part, however, it may be concluded that the electron microscope and the nitrogen surface area values obtained in the present work are in good agreement with those obtained by previous workers and tend to confirm the approximate correctness of both of these methods for obtaining, respectively, particle size distribution and surface areas.

[20] Private communication on work of Godfrey L. Cabot Company.
[21] B. S. Garvey and J. A. Freese, Jr., "A summary study of carbon black in GRS," issued privately by the B. F. Goodrich Company, Akron, Ohio (1943).

[22] Smith, Thornhill, and Bray, Ind. Eng. Chem. **33**, 1303 (1941).

One might raise the question as to whether some of the minor discrepancies that do exist for the particle diameters as obtained by the electron microscope method may not be due to an effect recently reported by Watson.[14] He claims that during the exposure of samples of some blacks to a stream of electrons in taking the electron microscope pictures, there is a tendency for the carbon particles to grow appreciably in size. In fact, he reports that the mean diameter of Micronex increased from 295 to 345A for "film" mountings, and from 295 to 495A for "filmless" mountings in three minutes of bombardment at normal focusing intensity. The size increase is presumably the result of the deposition of carbon from various organic sources on to the sample of black and is a function of a number of factors including the distance from the screen to the particular group of carbon black particles being photographed. There is no very good way of estimating the extent to which this effect may have entered into some of the results presented in Table II. However, the uniformity of the electron microscope values for a given black as obtained by various workers, and the general agreement between the surface areas, as measured by nitrogen and those that are obtained by calculating from d_A obtained from the electron microscope results, both seem to indicate that this particle growth effect has not been serious in previously published work on carbon black. In the present work, we feel very sure that no error of the type mentioned by Watson is involved. In the first place, the agreement between our own values of d_n and those reported by Watson for photographs so taken as to minimize or avoid particle growth is very striking and convincing. For P33, the value of d_n here reported is 1170A, compared to 1050A reported by Watson; for Shawinigan acetylene black, the d_n values are 445A in the present paper and 430A by Watson; and for Grade 6 carbon, the value here reported is 341A, compared to 295A obtained by Watson for a sample of Micronex W-6, which is similar to Grade 6 carbon. Furthermore, Dr. Hall,[13] who took the photographs used in the present work, was fully cognizant of the particle growth effects mentioned by Watson and took pains to avoid such growth in all his photographs. Accordingly, it seems certain that in this work the effect mentioned by Watson was not operative.

Briscoe and Warren[23] have concluded from their x-ray examination that carbon black particles are in reality clusters of packets or groups of graphite-like layers. These packets are in the range 12 to 60A in size and are, accordingly, smaller by an order of magnitude than the diameters calculated from the nitrogen adsorption isotherms for Grade 6, P33 acetylene black or lampblack. The question naturally arises as to how the existence of these small packets in clusters can be reconciled with the adsorption work. The answer seems to be that the packets in the clusters are so closely associated with each other that even nitrogen molecules cannot penetrate into the cluster. If this close piling together of the packets within a cluster is assumed, the nitrogen adsorption data are in good agreement with the concept of Briscoe and Warren as to the structural make-up of carbon black particles. In other words, electron micrographs, low angle scattering of x-rays and nitrogen adsorption isotherms all yield estimates of what Briscoe and Warren call "clusters" of the tiny packets and what most other workers designate as the "particle size" of carbon black; the x-ray structural measurements, on the other hand, yield the dimensions of the packets of graphite-like layers of which the carbon black particles are composed.

In conclusion, it should be made very clear that the electron microscope is an excellent apparatus for obtaining either particle size or size distribution in the range covered by the carbon blacks. On the other hand, it can yield surface areas only for those blacks that are non-porous and only within the uncertainty involved in assuming a shape factor and a value for the density of carbon. In contrast to this, the nitrogen adsorption method will yield apparently a reliable value for the surface area of either the porous or the non-porous blacks, but will give an average d_A value only for non-porous blacks, and will give no information at all in regard to the distribution of the size of the particles. Furthermore, of course, the calculation of the average diameter from the surface area measurements entails the same uncertainty as regards the density of the carbon as

[23] Briscoe and Warren, J. App. Phys. **13**, 364 (1942).

does the reverse calculation of area from the particle size measurements by the electron microscope. Accordingly, these two methods supplement rather than compete with each other. Each is superior to the other for certain purposes. Statements such as those made by Weigand[24] and by Weigand and Ladd[25] to the effect that the electron microscope is the only means of obtaining reliable surface areas are erroneous, as would be any statements that might be made relative to the superiority of the adsorption method over the electron microscope for obtaining particle size of finely divided substances.

[24] Wiegand, India Rubber World **105**, No. 3, 270; Can. Chem. Process Ind. **25**, 579 (1941).
[25] Wiegand and Ladd, Rubber Age **50**, 431 (1942).

ACKNOWLEDGMENT

It is a pleasure to express our appreciation to the Godfrey L. Cabot Company for financing this research, to Dr. Walter Smith of that company for his many helpful suggestions and for arranging to have electron microscope pictures made on the same samples on which we measured the nitrogen adsorption, and to Dr. C. E. Hall of the Massachusetts Institute of Technology for furnishing the electron microscope photographs together with a critical discussion of the technique he used in mounting the samples and taking the electron microscope photographs. We also appreciate being able to include some unpublished electron microscope results obtained by the B. F. Goodrich Company.[21]

The Visco-Elastic Behavior of a Highly Plasticized Nitrocellulose in Compression under Constant Load*

GEORGE J. DOYLE AND RICHARD M. BADGER
*Gates and Crellin Laboratories of Chemistry, California Institute of Technology, Pasadena, California***

(Received September 12, 1947)

The behavior of a highly plasticized nitrocellulose in compression under constant load has been investigated over a rather wide range of load and temperature. The load dependence appears to require the assumption of non-Newtonian viscosity. Though the observations are not quantitatively accounted for by the Tobolsky-Eyring model, the load and temperature dependence of the rate of compression appear to have some relation to the predictions of these authors.

INTRODUCTION

THE investigation here reported was a part of a program of investigation of the mechanical properties of smokeless powders recently conducted in this laboratory under a contract with the Bureau of Ordnance. This work included the study of plasticized materials in compression under constant load, and in large part consisted in the comparison of different materials. It was found desirable, however, to study at least one material under a rather wide variety of conditions of load and temperature. We believe the results obtained may be of general interest because of their relation to recently developed theories of visco-elastic behavior, and because the highly plasticized material studied is rather different from the materials on which somewhat similar studies have been reported.

MATERIAL STUDIED

The material studied was essentially a double base smokeless powder prepared by a solventless process. It contained 54.5 percent nitrocellulose and 43 percent nitroglycerine, the major plasticizing agent. The remainder of the material comprised stabilizer and inorganic materials in small amounts. The nitrocellulose was prepared from cotton linters and had a nitrogen content of about 12.6 percent. Its intrinsic viscosity in acetone solution was about 2.84, corresponding to a weight average degree of polymerization of about 390.[1]

* This paper is based on work done for the Bureau of Ordnance under contract NOrd-9652 with the California Institute of Technology.
** Contribution No. 1143.

[1] See Blaker, Badger, and Noyes, J. Phys. and Coll. Chem. **51**, 574 (1947).

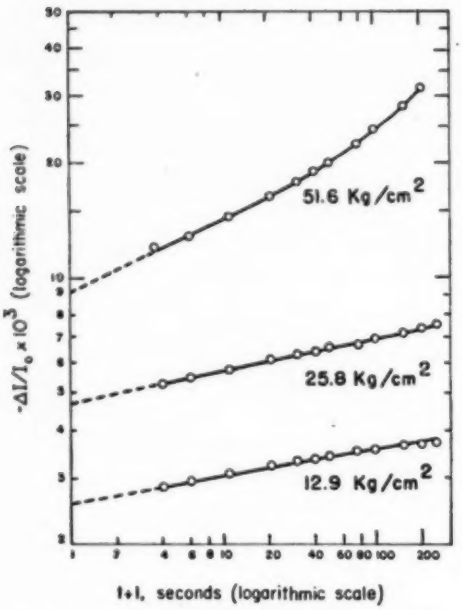


FIG. 1. Compression measurements at 20°C.

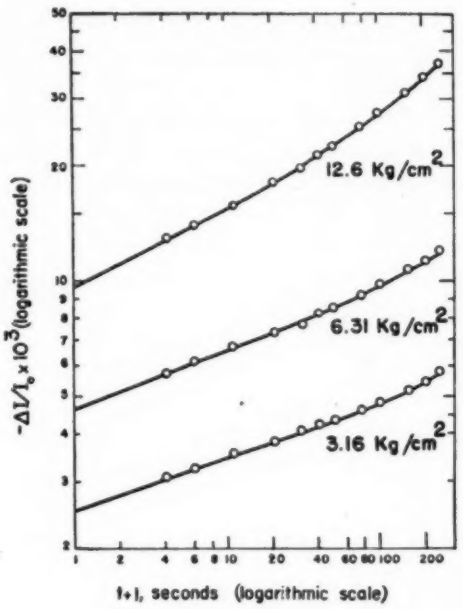


FIG. 2. Compression measurements at 50°C.

PREPARATION OF SPECIMENS

The material was extruded in $\frac{1}{4}$-inch strands which were annealed at about 60°C, while held in "V" grooves to produce straight specimens without strain. These were then aged for three months in stoppered tubes at room temperature. The test specimens, 0.35–0.40 inch in length, were then cut and were stored in an air thermostat at 20° and 50 percent relative humidity for five days. Before use, the ends of the specimens were made plane and parallel by dressing with fine emery paper, with the use of a jig. Considerable practice is required to do this satisfactorily. In our early work we were greatly troubled by "end effects" due to the failure of the specimens to seat properly on the anvils. In the studies here described this was almost entirely eliminated by exercising the greatest care.

The specimens to be used in a given series of measurements were stored in cylindrical, stoppered cavities in a massive brass block in which they fitted loosely. In this block they were brought to the temperature of the measurement. Temperature equilibration was found to be rapid, and when brought to any given temperature the specimens rapidly came to a steady state which did not alter during the course of several hours.

APPARATUS

The apparatus used was a specially designed instrument which could be used either for hardness tests or for compression measurements. In certain respects it resembled conventional hardness testers but differed from them in the greater range of compressions which could be measured and in the greater choice of loads which could be employed. When used for compression measurements, the instrument was provided with two circular anvils. The surfaces between which the specimen was placed were plane parallel to within 0.0002 inch. The upper anvil was fixed during an experiment. The lower one terminated in a plunger moving in a bearing and rested on one end of a balance arm. The relative motion of the two anvils was measured with a dial indicator reading 0.0001 inch per scale division. The load was applied to the specimen by hanging weights from either of two points on the longer end of the balance arm. Precision ball bearings were used, rather than knife edges, for all balance bearings. A somewhat complicated hydraulic device permitted weights to be applied or removed almost without shock at definite times.

The anvils of the compression balance were enclosed in an air thermostat in which a rapid circulation of air at 50 percent relative humidity

was provided. The temperature was kept constant within 0.5°C. Thermocouples were located at various points of the instrument to permit checking the uniformity of temperature and to monitor the adjustment of a heater and water jackets which controlled the flow of heat through the metal of the balance to the anvils in the thermostat.

EXPERIMENTAL PROCEDURE

Compression measurements were made at 20, 30, 40, and 50°C, with three different loads in each case which covered a range of four to one. At 25° four loads were used. In order to keep the compression within reasonable limits the loads employed had to be decreased at the higher temperatures, but they were so chosen that there was some overlapping of the ranges covered at two successive temperatures.

As mentioned above, the test specimens were approximately $\frac{1}{4}$ inch in diameter and $\frac{3}{8}$ inch in length. The diameter was determined by the material available, but considerable preliminary experimentation was done before the length chosen was found to offer the best compromise between two opposing difficulties. It was our purpose to study, so far as possible, the intrinsic properties of the plastic material rather than behavior largely conditioned by the shape of the specimens. Longer specimens exhibited unsatisfactory behavior under compression, which might perhaps be described as slender column failure although the ratio of length to cross section was much less than that at which such failure occurs with non-plastic materials. The compression of much shorter specimens could not be measured with adequate accuracy, and "end effects" were more serious.

When the compression of a specimen exceeded a few percent of its length, its form became noticeably "barrel shaped." The compressions studied were limited to the range in which this effect was not serious, though its consequences may have been entirely negligible.

Except at 25° all measurements were made on three specimens and the results reported are average values. To avoid uncertainties due to possible variability of composition the same specimens were used throughout at any given temperature. The measurements were made in the order of increasing load to minimize the effects of previous strain history. At least one-half hour was allowed for recovery between successive measurements on the same piece. Recovery was within 0.0002–0.0005 inch of the original length, except at the highest load and temperature where it was within 0.0008 inch. During the entire course of the measurements any permanent deformation of the specimens was so small as to be negligible.

The compression was followed by visual observation of the dial indicator. The first compression recorded was three seconds after the application of the load. Subsequent measurements were taken at suitable intervals, but terminated at 250 seconds.

EXPERIMENTAL RESULTS

Figures 1 and 2 show typical results of the compression measurements. They present the averages of three sets of measurements at 20° and at 50°C, respectively. Within the range of temperatures and loads investigated the relation between the fractional decrease in length and the time t (in seconds) is adequately represented by the equation

$$\log(-\Delta l/l_0) = m\log(\tau + t + \alpha t^3) + \log b. \quad (1)$$

The observations would have to be extended to considerably shorter times to evaluate the constant τ with any accuracy, but to secure a good fit with the existing data during the early stages

TABLE I. Constants of Eq. (1).

Temp. °C	Stress kg/cm²	m	Inst. fract. comp. b	$\alpha \times 10^5$	Y_0(kg/cm²) $\times 10^{-3}$
20	12.9	0.0696	0.00258	0.0	
	25.8	0.0855	0.00468	0.0	5.9
	51.6	0.192	0.00910	6.4	
25	7.67	0.065	0.00179	0.0	
	12.7	0.066	0.00278	0.0	5.2
	19.0	0.072	0.00401	0.0	
	25.6	0.096	0.00508	0.0	
30	7.67	0.0734	0.00232	0.0	
	19.2	0.0952	0.00520	2.4	3.9
	32.0	0.171	0.00851	6.5	
40	3.79	0.0937	0.00252	6.0 (?)	
	7.57	0.105	0.00432	2.8	1.9
	18.9	0.167	0.0104	5.8	
50	3.16	0.134	0.00253	3.2	
	6.31	0.150	0.00466	2.6	1.3
	12.6	0.208	0.0096	3.8	

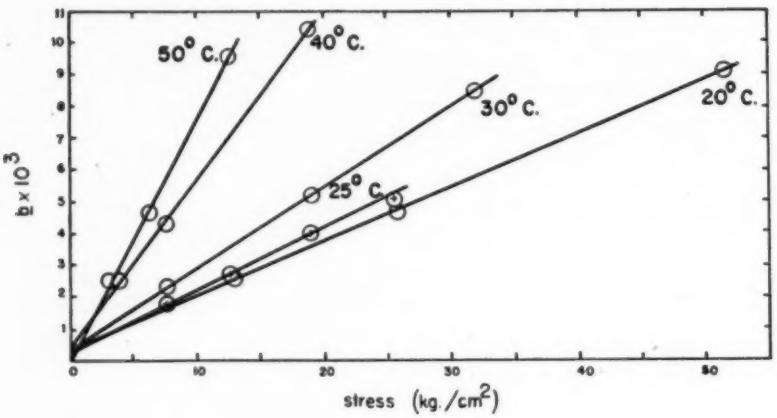


FIG. 3. Dependence of the constant b on stress at various temperatures.

of compression it seems necessary to give this constant a value different from zero. In plotting the data we have arbitrarily set $\tau = 1$ in all cases, and, as may be seen by reference to the figures, the fit is within the probable error. At moderate loads and temperatures the plots of $\log(-\Delta l/l_0)$ *versus* $\log(t+1)$ are straight lines within the time interval investigated. At higher loads and temperatures the plots show upward curvature which may be due in part to the deviation from cylindrical shape of the specimens when these are considerably compressed. This curvature may be taken into account in several ways. We have chosen to do so by the term in t^3 in Eq. (1).

Values of the constants m, α, and b fitting the data at the various loads and temperatures are given in Table I.

TABLE II. Fit of compression measurements with relation: $mF/A = a \sinh(cF/A)$.

Temp. °C	a (kg/cm²)	c (cm²/kg)	Stress, F/A (kg/cm²)	m (observed)	m (calc. from Eq. (4))	Deviation of calculated m percent
20	1.09	0.0565	12.9	0.0696	0.0673	−3.3
			25.8	0.0855	0.0860	+0.6
			51.6	0.192	0.195	+0.16
25	0.93	0.0643	7.67	0.065	0.062	−4.6
			12.7	0.066	0.067	+1.5
			19.0	0.072	0.076	+5.6
			25.6	0.096	0.091	−5.2
30	0.865	0.0790	7.67	0.0734	0.0726	−1.1
			19.2	0.0952	0.0979	+2.8
			32.0	0.171	0.168	−1.8
40	0.92	0.103	3.79	0.094	0.097	+3.2
			7.57	0.105	0.104	−1.0
			18.9	0.167	0.168	+0.6
50	0.90	0.144	3.16	0.134	0.133	−0.7
			6.31	0.150	0.143	−4.7
			12.6	0.208	0.214	+2.9

In giving a purely formal interpretation of Eq. (1), the constant b may be regarded as an "instantaneous" fractional compression which occurs immediately upon the application of the load. At any given temperature b increases linearly with the load, well within experimental error, as may be seen in Fig. 3. The failure of some of the plots to pass through the origin is believed to result from residual "end effects" previously mentioned. In previous work it was found that improvement in the technique of preparing specimens always reduced the zero-load intercept of these plots. From the slopes of the straight lines of Fig. 3 one may calculate a fictitious "instantaneous Young's modulus" by the relation:

$$Y_0(\text{kg/cm}^2) = 1 \Big/ \frac{db}{d(F/A)}, \qquad (2)$$

where F is the load in kilograms, and A the cross-sectional area of the specimen in square cm.

The constants b or Y_0 have a simple dependence on temperature which is very well expressed by the relation:

$$Y_0(\text{kg/cm}^2) = 1.34_8 \exp(-10{,}300/1.986T), \qquad (3)$$

where T is the absolute temperature.

The constant m of our empirical equation shows a rather complicated dependence on load and temperature. The available data are not adequate to determine uniquely this relationship, but the load dependence is represented, within

experimental error, by the equation:

$$mF/A = a \sinh(cF/A). \tag{4}$$

With the available data considerable latitude is permitted in the choice of the constants a and c, but in Table II pairs of constants are given at each temperature which represent the data with the accuracy shown in the final column of the table.

In the range investigated, the constant a does not change greatly with temperature. The temperature dependence of c is well represented by the relation:

$$c = \frac{1}{T} \exp\left(\frac{-6370}{1.986T} + 14.30\right),$$

as may be seen in Fig. 4.

DISCUSSION OF THE RESULTS

At any one load and temperature the behavior of our material, as represented in Eq. (1), could doubtless be simulated by a model consisting of spring and dashpot elements having an appropriate distribution of relaxation times;[2] we doubt, however, that the dependence of the rate of compression on load, which we have observed, could be accounted for in terms of a classical model without some rather arbitrary assumptions. This behavior appears to offer support for the non-Newtonian viscosity which has been assumed by Tobolsky and Eyring in their recent discussion of visco-elastic behavior.[3]

These authors have discussed a simple model, which, under compression during times which are not too long, should show behavior represented by an equation of the form:

$$-\Delta l = \beta[\ln(t+\delta) - \ln\delta], \tag{5}$$

where β and δ are constants at a given load and temperature. Equation (5) is inadequate to fit our data over more than very short time intervals. This is to be expected, since the model on which it is based comprises a single dashpot element. A further elaboration of this model along the lines already followed with classical models could probably be made to give satisfactory agreement with experiment, though the solution of the equations of motion presents mathematical difficulties.

At the present time it does not appear possible to attribute any very definite significance to the constants of our empirical equation, but we may point out a relation between our observations and the behavior of the Tobolsky-Eyring model which may have some significance. From Eq. (1), the rate of fractional compression at any time is given by:

$$\frac{d(-\Delta l/l_0)}{dt} = mb(\tau+t)^{m-1}, \tag{6}$$

if we ignore the term in t^3. At a time near the beginning of the compression, namely, when $(\tau+t)=1$, this rate is equal to mb or $mF/A\,Y_0$. The observed load dependence of this quantity is given above in Eq. (4). It is precisely the same as the load dependence of the initial rate of compression of the model of Eyring and Tobolsky, as

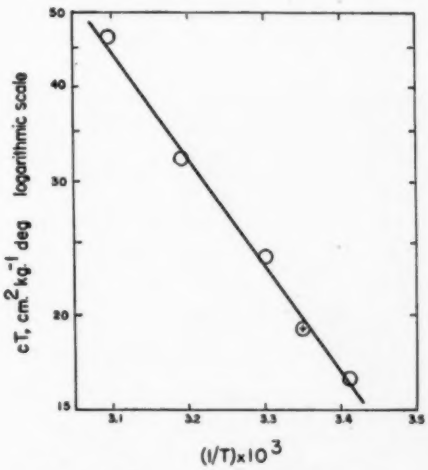


FIG. 4. Dependence of $\log cT$ on $1/T$.

may be seen by reference to their Eqs. (6).[3] The temperature dependence of mb or $mF/A\,Y_0$ is also very similar to that anticipated by these authors for their initial rate. To what extent this correlation has physical significance or is accidental can scarcely be decided until a more elaborate model has been discussed in detail.

[2] See, for example, R. Simha, J. Phys. Chem. **47**, 348 (1943).

[3] A. Tobolsky and H. Eyring, J. Chem. Phys. **11**, 125 (1943).

An Oxide Replica Technique for the Electron Microscope Examination of Stainless Steel and High Nickel Alloys*

E. M. MAHLA** AND N. A. NIELSEN**

(Received September 15, 1947)

An oxide replica method has been developed which is suitable for the electron microscope examination of the surfaces of stainless steel, nickel, and high nickel alloys. The oxide film is produced on the metal surface by oxidation in a molten nitrate solution and is chemically stripped from the base metal by a bromine-methanol solution. Replicas of this type have been used to study deep-etched structures, preferential etch attack, and secondary phases in austenitic and ferritic stainless steel and in nickel.

THE oxide-film method of preparation of replicas for use in examining opaque surfaces by electron microscopy has been developed to its greatest usefulness for the study of aluminum and aluminum alloys. Mahl[1] originated this oxide-film replica technique for aluminum. Keller and Geisler[2,3] have further developed and modified the method and studied precipitates in age-hardened aluminum alloys.

The general advantages of an oxide-film method are:

(1) The replica consists of a thin oxide film produced on and stripped from the prepared specimen. As such, it represents the actual surface layer of the metal.

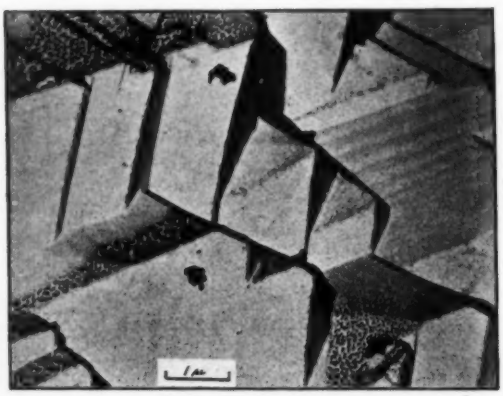


FIG. 1. Deep-etched 18-8-S-Mo stainless steel. (Bromine-methanol etch. Replica was shadowed with gold.) 8500×.

(2) Fewer artifacts are formed in this type of replica than in mechanically stripped replica films.

(3) Excellent contrast and resolution can be obtained. Variations in surface composition resulting from the presence of secondary phases, precipitates, etc. affect the electron-scattering ability of the film and thereby result in improved image contrast and definition.

The prime requirement of oxide replicas is that they should themselves be essentially "structureless." Oxide films that are granular or porous or that possess considerable "fine structure" cannot adequately replicate the structural detail of the base metal.

Oxide films from metals other than aluminum have been examined by electron microscopy. Gulbransen, Phelps, and Hickman[4,5] have studied crystal size, shape, distribution, etc. in oxide films electrolytically isolated from metallographically

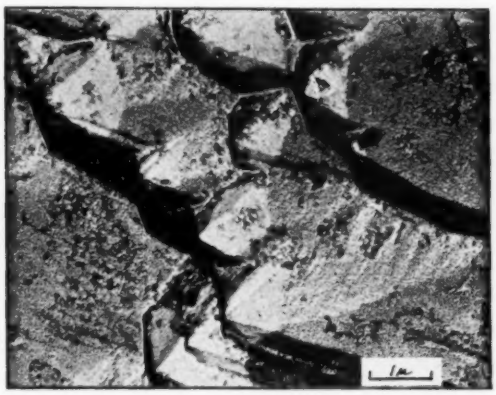


FIG. 2. Deep-etched 18-8-S-Mo stainless steel. (Bromine-methanol etch. Replica was shadowed with gold.) 8500×.

* Presented at a meeting of the Electron Microscope Society on December 11, 1947.

** Engineering Research Laboratory, Engineering Department, E. I. du Pont de Nemours and Company, Wilmington, Delaware.

[1] H. Mahl, Zeits. tech. Physik **21**, 17 (1940).

[2] F. Keller and A. H. Geisler, Trans. Am. Inst. Mining Met. Eng. **156**, 62 (1944).

[3] F. Keller and A. H. Geisler, Metals Tech. T.P. No. 2108 (October, 1946).

[4] E. A. Gulbransen, R. T. Phelps, and J. W. Hickman, Ind. Eng. Chem. (Anal. Ed.) **18**, 391 (1946).

[5] E. A. Gulbransen, R. T. Phelps, and J. W. Hickman, Ind. Eng. Chem. (Anal. Ed.) **18**, 640 (1946).

polished specimens of numerous metals and alloys oxidized at high temperature in 0.1 atmos. of oxygen. Within the experience of the present authors, air oxidation does not generally produce "structureless" oxide films (at the magnification level of the electron microscope) but results in a porous or non-uniform type of film. Bennek *et al.*[6] have described an aluminum replica method for steel and examined oxide films produced on steel by concentrated chromate solutions.

PREPARATION OF OXIDE REPLICAS

Although no extensive investigation of oxidation procedures has been made, the authors have found that heat-tinting specimens of stainless steel, nickel, or high nickel alloys in a molten mixture of $NaNO_3-KNO_3$ (equal parts by weight) produces oxide films that, when stripped, serve as satisfactory surface replicas. The metal specimen is prepared by usual metallographic techniques, or it may simply be deep-etched with a suitable corrosive agent. The specimen is then immersed in the molten $NaNO_3-KNO_3$ until it is superficially oxidized to the extent that the surface shows a brown interference tint. A considerable degree of control of the oxidation conditions is possible, and visual inspection of the sample is sufficient to determine when the sample should be removed from the salt bath. For instance, a sample of deep-etched austenitic stainless steel was oxidized for 5 min. at 800°F (425°C) to produce a satisfactory oxide replica film. However, there is considerable latitude in the time and temperature of oxidation, and an exact specification is not critical.

After oxidation, the specimen is cooled in air, washed with distilled water, and dried. After $\frac{1}{8}$-in. squares have been scribed with a razor blade or a sharp-pointed instrument in the oxidized surface, the specimen is wholly immersed in a solution of bromine in methanol. The concentration of this solution is not critical. Nickel specimens require rather dilute solutions, 2–3 ml of bromine per 100 ml of methanol, while a stronger concentration, 7–10 ml of bromine in the 100 ml of methanol, is necessary for stainless steel specimens. The bromine solution attacks the film-free metal exposed along the scratches and effectively "undermines" the oxide film, freeing it from the metal. When the film is completely stripped (a matter of several minutes to several hours according to the metal composition) small squares of oxide film float free or cling loosely to the base metal. Gentle agitation of the solution or specimen serves to detach this film. The squares of oxide film are collected with a piece of 200-mesh wire screening and are washed by transfer through successive dishes of fresh methanol. The individual squares of film are then caught on a microscope screen, dried, and examined in the electron microscope.

If a particular area on the metal specimen is to be examined, this area only is originally scribed in the metal surface. The one square of film is easily isolated for investigation.

FIG. 3. Deep-etched 18-8-S-Mo stainless steel. (Bromine-methanol etch.) 8500×.

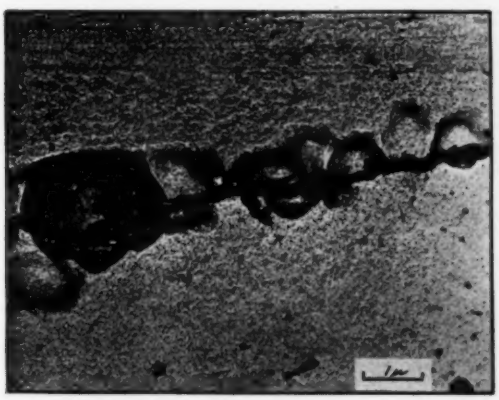


FIG. 4. Grain boundary in sensitized 18-8-S-Mo stainless steel. (Etched electrolytically with 10 percent oxalic acid after metallographic polish.) 8100×.

[6] H. Bennek, O. Rüdiger, F. Stäblein, and K. E. Volk, Arch. Eisenhüttenw. **15**, 431 (1942).

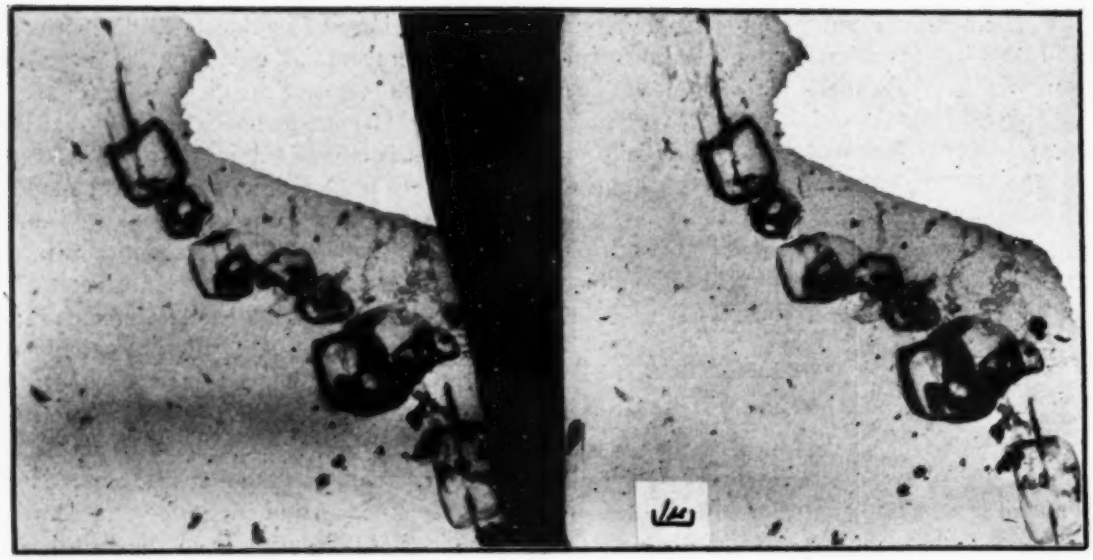


FIG. 5. Stereograph of grain boundary in sensitized 18-8-S-Mo stainless steel. (Metallographic polish, electrolytic oxalic acid etch.) 4800×.

ELECTRON MICROGRAPHS OF OXIDE REPLICAS

The accompanying figures are examples of electron micrographs of surface oxide replicas prepared by the above technique. Figures 1–3 illustrate a deep-etched structure (bromine-methanol etch) found in 18-8-S-Mo stainless steel. The action of the etchant has been to corrode preferentially the (111) planes. This is most clearly shown in Fig. 1. In Fig. 2 small octahedra formed as a result of this preferential corrosion mechanism are to be seen in relief from the plane of the base metal. Figure 3 shows the junction of two (111) planes. "Etch-steps" have formed in these planes and the planes themselves are minutely etched.

(Note: Before examination in the electron microscope the replicas of Figs. 1 and 2 were "gold-shadowed." The effect of the shadowing was only to "spot" certain crystalline

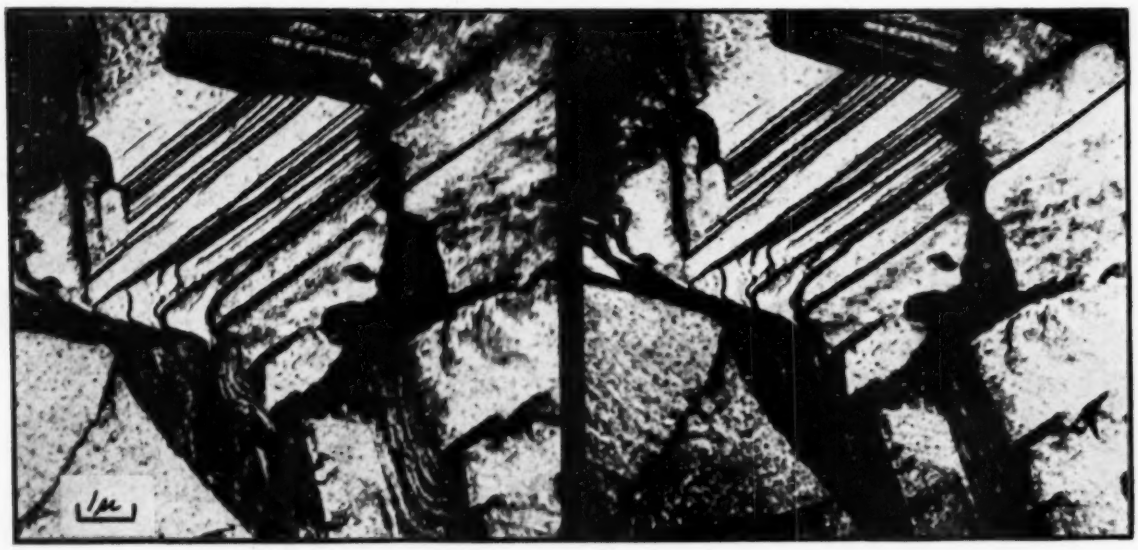


FIG. 6. Stereograph of deep-etch structure in 18-8-S-Mo stainless steel. (Bromine-methanol etch.) 7500×.

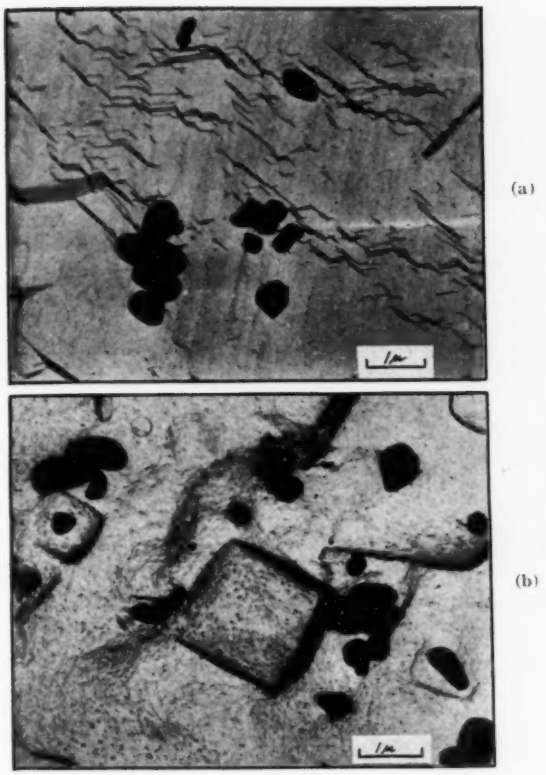


FIG. 7. Deep-etched 15-16 Cr stainless steel. (a) Top: Bromine-methanol etch. 7750×. (b) Bottom: Sulfuric acid etch. 8500×.

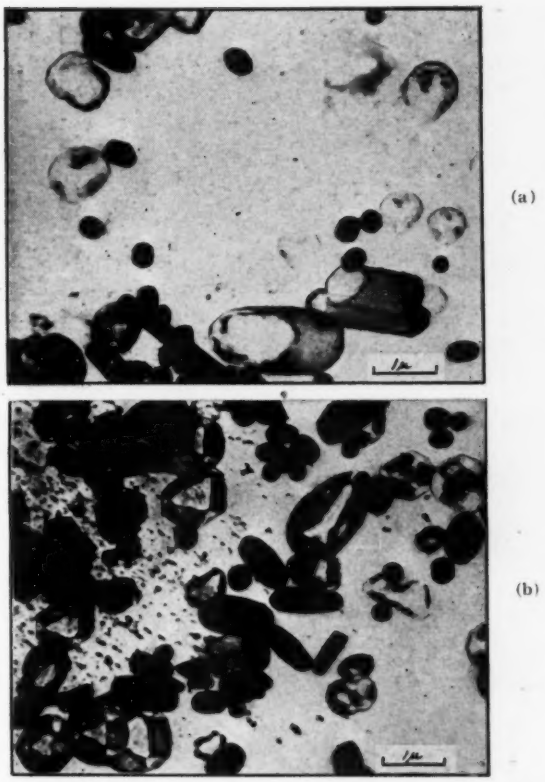


FIG. 8. Structure of 27 Cr stainless steel. (a) Top: Metallographic polish, Vilella's etch. 8500×. (b) Bottom: Same. 8500×.

facets which were normal (or nearly so) to the path of evaporation. Metal shadowing did not enhance the resolution and image contrast.)

Figure 4 is an oxide replica of a polished and lightly etched specimen of sensitized 18-8-S-Mo stainless steel. What are commonly termed "grain-boundary carbides" are in this instance only holes along an etched-out grain boundary. The effect of holes is more clearly evident when stereographs are examined (Fig. 5). Stereographs have also been effectively used on deep-etched structures (Fig. 6). The same effects have been obtained by preparation of Polaroid vectographs of these stereomicrographs.

Chemical stripping of oxide films from ferritic stainless steels is accompanied by the production of gray-black powder, resulting in contamination of the oxide film. Even after careful washing in methanol, particles of this powder still adhere to the replica. Figure 7a is an electron micrograph of a bromine-methanol-etched surface of 15–16 percent Cr stainless steel. Some development of crystalline facets can be seen. The oxide film itself appears perfectly structureless. In Fig. 7b the corrosive attack of sulfuric acid on the same alloy has resulted in development of some cubic faces. Figures 8a and 8b are micrographs of 27

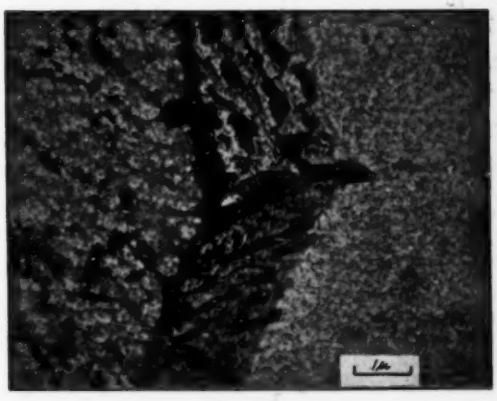
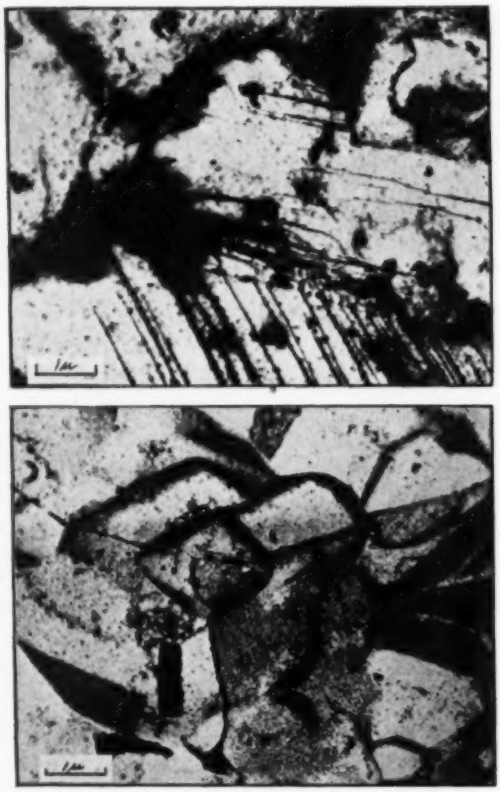


FIG. 9. Sigma and austenite in 25-12-Cb stainless steel. (Metallographic polish. Dilute nitric-hydrochloric acid etch.) 7750×.



(a)

(b)

FIG. 10. Deep-etched nickel. (a) Top: Bromine-methanol etch. 8500×. (b) Bottom: Nitric acid-acetic acid etch. 8500×.

percent Cr stainless steel. The film is heavily contaminated with the dense "corrosion product powder," but particles believed to be carbides can be seen interspersed. The carbides apparently have suffered varying degrees of etching; those in Fig. 8b show what appear to be prism faces.

Figure 9 was made from an oxide replica of a specimen of 25-12-Cb stainless steel containing "islands" of sigma-phase. The micrograph shows the junction between sigma and austenite and indicates that the metallographic etchant produced very severe corrosion at the junction of the two phases. The sigma apparently corroded preferentially to the austenite.

The effects of different etching treatments are again illustrated by the micrographs of oxide replicas from nickel surfaces. A bromine-methanol etch (Fig. 10a) developed a closely spaced series of "etch steps." A nitric-acetic acid etch (Fig. 10b) produced a surface showing none of these effects, but instead a rather rounded, hilly topography. This type of structure might more logically be expected from a face-centered cubic structure such as nickel.

The above figures show the extent to which this type of oxide replica has been investigated by the authors. The work has not included a direct comparison with other techniques, for instance, the silica replica method. The present information is being presented in the hope that it will be of use to other investigators.

CONCLUSIONS

1. An oxide replica technique that is applicable to the stainless steels, nickel, and high nickel alloys has been developed.

2. The technique has been shown to be useful in the study of deep-etch structures, preferential corrosion mechanisms, and etching attack. Secondary phases such as carbides, sigma-phase, etc. are effectively reproduced for study by this type of replica.

ACKNOWLEDGMENT

The electron micrographs were taken by R. H. Fisher and J. F. Godlewski under the supervision of C. E. Willoughby of the Chemical Department.

The Head-On Collision of a Shock Wave and a Rarefaction Wave in One Dimension

H. E. Moses
*Department of Engineering Research, University of Michigan, Ann Arbor, Michigan**
(Received October 20, 1947)

A procedure for investigating the strengthening of a shock which collides head-on with a rarefaction wave is suggested and is carried through for the case in which the entropy jump across the shock is small enough to be negligible.

I. INTRODUCTION

THE problem which is considered in the present paper is the strengthening and acceleration of a shock which moves in such a way that it collides with a rarefaction which approaches the shock from the low pressure side. Such a collision may be thought to take place in a semi-infinite tube as shown in Fig. 1. The x axis is taken to be the axis of the tube. A piston is initially moving toward the left with a constant velocity less than that of sound and the gas in the tube moves with the same velocity. (The restriction on the initial velocity of the piston is not essential; it is merely convenient for the purpose of describing the phenomenon.) When the piston reaches the position $x=0$ at time $t=0$, it is assumed that the piston accelerates in some manner toward the left and thus produces a simple rarefaction wave which moves toward the right. At a large positive value of x a shock wave whose low pressure side faces the origin is assumed to exist. The shock wave will move toward the origin with uniform speed and constant strength until it meets the oncoming rarefaction wave at a distance L from the origin. The shock will then accelerate and be strengthened as it meets the gas of decreasing density. Furthermore, the entropy jump across the shock, which is constant as long as the shock moves through the gas which has not yet been disturbed by the rarefaction, will increase. Thus an entropy wave which moves with the fluid will be formed on the high pressure side of the shock so that the flow on the high pressure side of the shock is no longer isentropic.

In reference 1 Courant and Friedrichs consider the interaction discussed above and conclude that the final result of such an interaction will be a shock wave moving towards the left and a rarefaction wave moving toward the right, separated by a zone of gas of varying entropy.[1] The calculations in the present paper have been made with these results kept in mind and will describe in more detail than is given in reference 1 the actual process of interaction between the shock and rarefaction waves.

It might be pointed out that the collision between the shock wave and rarefaction wave as discussed in the present paper is an idealization of processes which occur in intermittent jet engines and supersonic wind tunnels which operate by permitting air from the outside atmosphere to pass through the tunnel into a low pressure reservoir.

II. PROCEDURE

Before discussing the interaction between the rarefaction and shock, the properties of the rarefaction and shock waves will be reviewed briefly. The terminology of reference 2 will be used throughout.[2]

A simple wave in non-steady, one-dimensional flow refers to a special isentropic flow in which the fluid velocity, pressure, density, and speed of sound assume constant values along each straight line of a one-parameter family of straight lines in the $x-t$ plane. These values in general differ from line to line of this family. For general non-steady one-dimensional flows, two families of curves in the $x-t$ plane play particularly important roles. These curves are called the characteristic curves and are defined by the differential

* Now with Department of Physics, Columbia University.

[1] R. Courant and K. Friedrichs, "Interaction of shock and rarefaction waves in one-dimensional motion," OSRD Report No. 1567.

[2] "Supersonic flow and shock waves (shock wave manual)," AMP Report 38.2R.

equations $dx/dt=u\pm c$, where u is the velocity of the fluid and c is the local velocity of sound. When the flow is a simple wave, one family of these characteristics is the family of straight lines discussed above. A fundamental theorem on simple waves says that flows adjacent to flows of steady state are simple waves. Hence, the rarefaction wave described above, produced by acceleration of the piston, is a simple wave, since the fluid is initially in a steady state. In the present case the family of characteristics which are straight lines is that one which has the plus sign.

In a simple wave the pressure p, the density ρ, and the velocity of sound c are related to the velocity of the fluid by the following formulas:

$$p=p_i\left[1+\frac{\gamma-1}{2}\frac{u-u_i}{c_i}\right]^{2\gamma/(\gamma-1)}, \tag{1}$$

$$\rho=\rho_i\left[1+\frac{\gamma-1}{2}\frac{u-u_i}{c_i}\right]^{2/(\gamma-1)}, \tag{2}$$

$$c=\left(\frac{\gamma p}{\rho}\right)^{\frac{1}{2}}=c_i\left[1+\frac{\gamma-1}{2}\frac{u-u_i}{c_i}\right], \tag{3}$$

where subscript i refers to the initial constant state of the gas before the rarefaction has affected the flow and γ is the ratio of specific heats.

The properties of shock waves will now be discussed. Let the subscripts 0 and 1 refer, respectively, to the low and high pressure sides of the shock. There are three shock conditions which arise from the conditions of conservation of mass, momentum, and energy across the shock. These relations can be written

$$\rho_0^2v_0^2=\rho_0\rho_1\frac{p_1-p_0}{\rho_1-\rho_0}=\rho_1^2v_1^2, \tag{4}$$

$$\rho_0u_0v_0+p_0=\rho_1u_1v_1+p_1, \tag{5}$$

$$\rho_1=\rho_0\frac{\mu^2p_0+p_1}{\mu^2p_1+p_0}. \tag{6}$$

In the above equations

$$v_0=u_0-\dot{x}_1, \tag{7}$$

$$v_1=u_1-\dot{x}_1, \tag{8}$$

where x_1 indicates the position of the shock and the dot indicates differentiation with respect to time, so that $\dot{x}_1$ is the velocity of the shock. The constant μ^2 equals $(\gamma-1)/(\gamma+1)$.

It will be convenient to measure the shock strength in terms of the excess pressure ratio $\xi=(p_1-p_0)/(p_0)$. In terms of ξ the shock conditions may be written

$$v_0=c_0(1+\nu\xi)^{\frac{1}{2}}, \quad (\nu=1/(1+\mu^2)), \tag{9}$$

$$\xi=(\gamma v_0/c_0^2)(u_0-u_1), \tag{10}$$

and from (9) and (10)

$$v_0=\frac{-\nu\gamma(u_1-u_0)}{2}+\left(\frac{\nu^2\gamma^2(u_1-u_0)^2}{4}+c_0^2\right)^{\frac{1}{2}}, \tag{11}$$

$$\dot{x}_1=u_0+\frac{\nu\gamma(u_1-u_0)}{2}-\left(\frac{\nu^2\gamma^2(u_1-u_0)^2}{4}+c_0^2\right)^{\frac{1}{2}}, \tag{11a}$$

$$u_1-u_0=\frac{-c_0\xi}{\gamma(1+\nu\xi)^{\frac{1}{2}}}, \tag{12}$$

$$c_1=c_0\left(\frac{(1+\xi)(1+\nu\mu^2\xi)}{(1+\nu\xi)}\right)^{\frac{1}{2}}. \tag{13}$$

The process of interaction will now be considered. The quantities on the low and high pressure sides of the shock before the interaction will be denoted by the subscripts A and B, respectively. Thus, before the interaction of the shock wave and rarefaction wave $p_1=p_B$, $u_1=u_B$, $\rho_1=\rho_B$, $c_1=c_B$, $p_0=p_A$, $u_0=u_A$, $\rho_0=\rho_A$, $c_0=c_A$. Using this notation Eqs. (1), (2), and (3) may be rewritten as follows:

$$p_0=p_A\left[1+\frac{\gamma-1}{2}\frac{u_0-u_A}{c_A}\right]^{2\gamma/(\gamma-1)}, \tag{1a}$$

$$\rho_0=\rho_A\left[1+\frac{\gamma-1}{2}\frac{u_0-u_A}{c_A}\right]^{2/(\gamma-1)}, \tag{2a}$$

$$c_0=c_A\left[1+\frac{\gamma-1}{2}\frac{u_0-u_A}{c_A}\right]. \tag{3a}$$

Since in the wave $u_0<u_A$ (u_0 and u_A are negative in our coordinates), it is seen that $p_0<p_A$, $\rho_0<\rho_A$,

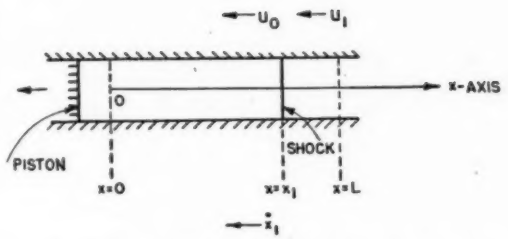


FIG. 1. Collision between shock and rarefaction wave.

$c_0 < c_A$. Furthermore, the head of the wave travels with velocity $u_A + c_A$.

Furthermore, if ξ_1 denotes the excess pressure ratio of the shock before interaction, then

$$u_B - u_A = \frac{-c_A \xi_1}{\gamma(1+\nu\xi_1)^{\frac{1}{2}}}. \tag{12a}$$

The rarefaction wave is completely determined by the velocity with which the piston is withdrawn from the tube. It will, therefore, be assumed that u_0, hence also p_0, ρ_0, c_0 are known functions of x and t.

It is the problem of the present paper to find x_1, $\dot{x}_1$, u_1, ξ_1 as a function of time as the shock wave moves along its path. If u_1 is known as a function of u_0, then the differential Eq. (11a) can be solved to give $x_1(t)$ and $\dot{x}_1(t)$. Then also $\xi(t)$ can be found from (10). Therefore, in addition to Eq. (11), another relation between v_0, u_0, u_1 is needed so that v_0, u_1 can be solved in terms of u_0 alone. Having found u_1 as a function of u_1 the procedures outlined above can be used to find the desired quantities.

As explained previously, the fluid on the high pressure side of the shock wave is *not* isentropic. Consider the shock at a given time. The region on the high pressure side can be divided into small regions in which the pressure, density, velocity, and entropy are considered constant. In particular consider the small region immediately adjacent to the shock. The fluid particles which have just passed through the shock will move toward the left with a velocity greater (i.e., more negative in the coordinate system chosen) than the fluid in the small region being considered. Hence the particles which have just passed through the shock may be considered the front of a rarefaction wave which will pass through the small region. Inasmuch as the fluid is isentropic in the region considered, Eq. (1) describes the relation between pressure and velocity.

In Eq. (1) we shall write

$$\begin{aligned} &p = p_1, \quad u = u_1, \quad c = c_1, \\ &u_i = u_1 + \Delta u_1, \quad c_i = c_1 + \Delta c_1, \\ &p_i = p_1 + \Delta p_1, \end{aligned} \tag{14}$$

passing to the limit, the following differential equation is obtained:

$$dp_1/du_1 = \gamma p_1/c_1. \tag{15}$$

By means of Eqs. (12), (13), and (1a), Eq. (15) can be converted into a differential equation with ξ and u_0 as variables. By integrating this differential equation, a new relation which gives ξ as a function of u_0 is obtained. By substituting for ξ in (12), $u_1 - u_0$ is found as a function of u_0 alone, and the procedure outlined above may be used to find $x_1(t)$, $\dot{x}_1(t)$, $\xi(t)$, and finally $u_1(t)$.

A particularly simple and interesting case is the one in which the shock is weak. As shown in reference 2, the entropy jump across a sufficiently weak shock is proportional to ξ^3. Thus if the shock is so weak that all powers of ξ higher than the second can be neglected the fluid on the high pressure side of the shock can be considered isentropic.

When the value of c_1 corresponding to isentropic flow is substituted in (15), we obtain

$$p_1 = p_B\left[1 + \frac{\gamma-1}{2}\frac{u_1-u_B}{c_B}\right]^{2\gamma/(\gamma-1)}. \tag{1b}$$

This result was to be expected from the manner of derivation of (15). It is also to be expected from the fact that we have a non-uniform isentropic state adjacent to the constant state given by p_B, u_B, ρ_B, therefore the wave on the high pressure side of the shock must be a simple wave and the simple wave relations must hold.

From (1b)

$$\xi = -1 + (1+\xi_1)\frac{\left[1 + \frac{\gamma-1}{2}\frac{u_1-u_B}{c_B}\right]^{2\gamma/(\gamma-1)}}{\left[1 + \frac{\gamma-1}{2}\frac{u_0-u_A}{c_A}\right]^{2\gamma/(\gamma-1)}},$$

and (13) may be replaced by its isentropic

counterpart

$$c_0/c_1=(1+\xi)^{(\gamma-1)/2\gamma}. \tag{17}$$

From (17)

$$c_A/c_B=(1+\xi_1)^{(\gamma-1)/2\gamma}. \tag{17a}$$

For the special case of weak shocks, which is to be treated, the general procedure outlined above will be modified as follows: From the nature of the problem it is expected that u_1 can be expanded in powers of ξ_1.

$$u_1(\xi_1, u_0)=f_0(u_0)+\xi_1 f_1(u_0)+\xi_1^2 f_2(u_0)+\cdots. \tag{18}$$

The functions f_0, f_1, $\cdots$ etc. will be sought. Using (18), (17a), and (12a) in (16), ξ is obtained as a power series in ξ_1 which involves the f_i's as coefficients. This series is not valid beyond the second power of ξ_1 because of the assumption of isentropic flow. The expression for ξ so obtained is substituted in Eq. (9) to obtain v_0 in an expansion in ξ_1, which again will not be valid beyond the second power of ξ_1. By substituting u_1 as given by Eq. (18) into Eq. (11) an alternative expansion of v_0 in powers of ξ_1 is obtained. Comparison of the coefficients of powers of ξ_1 of both expansions yields the functions f_0, f_1, and f_2.

Having found these functions, u_1 is known as a function of u_0; and x_1, $\dot{x}_1$, ξ, u_1, u_0 can be found as functions of time as the shock moves along its path, as explained previously.

III. RESULTS

A. Infinitesimal Shocks

For initially infinitesimal shocks $\xi_1=0$. It is easily verified that $f_0=u_0$ and that $\xi=0$. Therefore, an infinitesimal shock remains an infinitesimal shock when passing through a rarefaction wave.

Furthermore, $\dot{x}_1=u_0-c_0$, which is the equation of a backward characteristic of the simple wave. Hence, infinitesimal shocks move along the characteristics of the simple wave. This result was to have been expected from the role played by characteristics as propagators of small disturbances.

B. Weak Shocks

Weak shocks are defined as those shocks for which the powers of ξ higher than the first can be neglected.

It is found that $f_1=-c_A/\gamma$. Furthermore, to this approximation

$$u_1-u_B=u_0-u_A, \tag{19}$$

which shows that for weak shocks the increase in the velocity of the fluid behind the shock is equal to the increase in velocity which occurs ahead of the shock, or to put the result in different words, the velocity of the fluid behind the shock differs from the velocity ahead of the shock only by a constant.

The differential equation for the position of the shock is

$$\dot{x}_1=u_0+(\nu\gamma/2)(u_B-u_A)-c_0. \tag{20}$$

The path followed by a weak shock is, therefore, no longer a characteristic of the rarefaction wave. Moreover, since u_B-u_A is negative, the shock wave will travel faster than the velocity of an infinitesimal disturbance (i.e., a sound wave) through the rarefaction region.

The strength of the shock as measured by its excess pressure ratio increases as it passes through the rarefaction region. The expression for ξ is

$$\xi=\xi_1(c_A/c_0). \tag{21}$$

From (3a) it is clear that c_0/c_A decreases so that ξ increases. An initially weak shock may, therefore, become strong when interacting with the rarefaction.

In order to present a specific example of the method of finding the velocity of shock and of the fluid before and behind the shock, a special rarefaction wave will be considered, namely, the rarefaction wave which results when the piston undergoes infinite acceleration in changing its initial velocity u_A to some final constant velocity. Such a rarefaction wave is called a centered rarefaction wave. The velocity u_0 is given by

$$u_0-u_A=(1-\mu^2)[(x/t)-c_A-u_A]. \tag{22}$$

For simplicity, c_A will be taken equal to unity, u_A will be taken equal to zero, and the distance from the origin to L at which the shock and rarefaction interact will also be taken as unity. Likewise the time interval required by the head of the rarefaction wave to move from the origin to the position $x=L$ will be taken as unity. These simplifications correspond merely to a choice of units and a frame of reference from which the phenomenon is viewed.

Then for $t>1$

$$x_1=t\left[\frac{4(1-\mu^2)-u_B}{4\mu^2(1-\mu^2)}t^{-2\mu^2}-\frac{4(1-\mu^2)^2-u_B}{4\mu^2(1-\mu^2)}\right], \quad (23)$$

$$\dot{x}_1=\left[\frac{4(1-\mu^2)(1-2\mu^2)-u_B(1-2\mu^2)}{4\mu^2(1-\mu^2)}t^{-2\mu^2}-\frac{4(1-\mu^2)^2-u_B}{4\mu^2(1-\mu^2)}\right], \quad (24)$$

$$u_0(t)=(1-\mu^2)\left[\frac{4(1-\mu^2)-u_B}{4\mu^2(1-\mu^2)}t^{-2\mu^2}-\frac{4(1-\mu^2)^2-u_B}{4\mu^2(1-\mu^2)}-1\right], \quad (25)$$

$$u_1(t)=u_0(t)+u_B, \quad (26)$$

and the various quantities which are desired are completely solved for.

C. Moderately Strong Shocks

A moderately strong shock is one in which all powers of higher than the second can be neglected. For such shocks it is found that

$$f_2=(\nu/2\gamma)c_A. \quad (27)$$

Therefore, $u_1-u_B=u_0-u_A$ as for weak shocks. This result is somewhat surprising, inasmuch as one might expect the relation between u_1 and u_0 to be non-linear for moderately strong shocks. It therefore appears from this point of view that instead of using the excess pressure ratio as a criterion of shock strength it would be more satisfactory to use the non-dimensional velocity difference $(u_1-u_0)/c_0$. Thus, weak disturbances might better be defined as those such that all powers of $(u_1-u_0)/c_0$ beyond the first may be neglected. However, we shall continue to use our original definition.

For moderately strong shocks the excess pressure ratio is given by

$$\xi=\xi_1\frac{c_A}{c_0}\left[1+\frac{\gamma+1}{4\gamma}\xi_1\left(\frac{c_A}{c_0}-1\right)\right]. \quad (28)$$

Since $c_A/c_0>1$, the strength of the moderately strong shock increases at a faster rate than that of a weak shock in terms of c_A/c_0. Also the equation of motion of the shock is

$$\dot{x}_1=u_0+\frac{\nu\gamma}{2}(u_B-u_A)-c_0-\frac{\gamma^2\nu^2(u_B-u_A)^2}{8c_0}, \quad (29)$$

which shows that the moderately strong shock travels faster through the rarefaction region than the weak shock.

As before, when u_0 is given as a function of x and t, (29) may be integrated to give the position and velocity of the shock as a function of the time and then x_1, $\dot{x}_1$, u_0, u_1, ξ may be found as functions of the time or position of the shock.

IV. ACKNOWLEDGMENTS

The writer wishes to thank Professor A. M. Kuethe and other members of the Department of Aeronautical Engineering of the University of Michigan for their helpful advice.

The present work was done with funds supplied under a contract between the Air Forces and the Department of Engineering Research of the University of Michigan. The present report will appear as an External Memorandum of the Department of Engineering Research.

Preferred Orientation Determination Using a Geiger Counter X-Ray Diffraction Goniometer

B. F. Decker, E. T. Asp, and D. Harker
General Electric Research Laboratory, Schenectady, New York
(Received October 30, 1947)

A simple and accurate method of determining pole figures by using a special sample mount with a Geiger counter x-ray diffraction goniometer is described and examples given. Also, a correction formula is derived which takes into account absorption change and change in diffracting volume as the sample changes position with respect to the x-ray beam.

A METHOD has been devised to determine pole figure data by use of a Geiger counter x-ray diffraction goniometer. It has three important advantages over the photographic method: (1) greater speed, (2) more accuracy, and (3) elimination of the pole figure chart. When only one quadrant is needed to describe a preferred orientation, all data may be taken and the pole figure constructed in approximately two hours.

A special sample holder of simple design is required (see Fig. 1). The base of the holder is circular with degrees marked. A shaft which may be rotated about its axis fits in the center of the base. A flange fastened to the shaft has a scribe mark on it which passes over the degree marks on the base as the shaft turns, enabling one to read the amount of rotation of the shaft. Mounted in a circular groove on the top of the shaft is a vertical ring which may be turned in the groove, thus giving rotation about the axis of the ring. The ring is marked in degrees, and a scribe mark on the edge of the groove permits one to read the amount of rotation of the ring. The sample is mounted in the center of the ring (using Scotch tape, plasticene, etc.). The entire sample holder fits on the machine in the same place as the normal specimen holder so that the center of the vertical ring is in the direct x-ray beam and at the center of the spectrometer arc. (Fig. 2.) Also, the sample holder moves through an angle θ as the Geiger counter moves through 2θ. Thus, the angle scale on the base may be inscribed so that the zero setting of the shaft will bring the vertical ring into position to bisect the angle between incident and diffracted rays for any value of θ.

In operation, the Geiger counter is moved into position to receive the diffracted beam from the family of planes being studied. When the shaft is in its zero position, the intensity measured by the Geiger counter is due to diffraction by vertical planes normal to the plane of the vertical ring, whose poles, therefore, lie in directions parallel to the horizontal diameter of the vertical ring. (Fig. 3.) If the plane of the pole figure is taken to be the plane of the vertical ring this intensity is recorded at each end of the horizontal diameter of the pole figure. Now, let the shaft be moved from its zero position n degrees clockwise as viewed from above. In this position the intensity measured is the result of diffraction by vertical planes located n degrees counterclockwise from

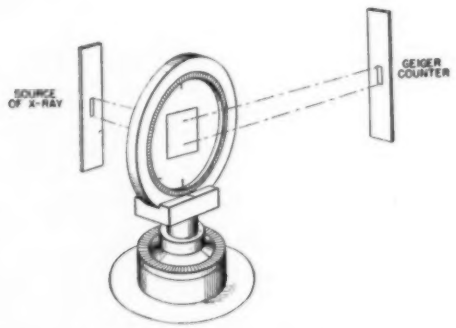


Fig. 1. Special sample mount used for preferred orientation determinations.

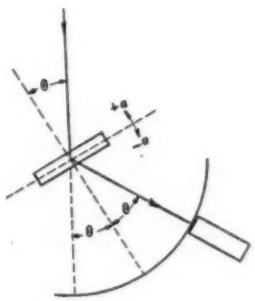


Fig. 2.

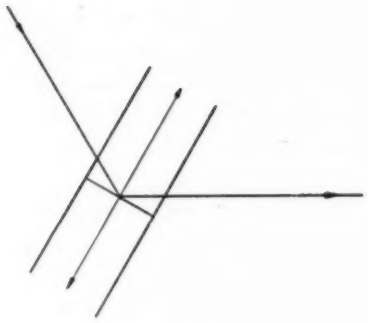


FIG. 3.

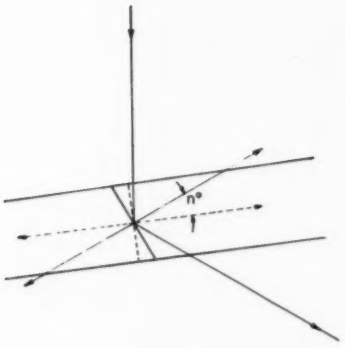


FIG. 4.

the vertical planes normal to the surface of the specimen. (Fig. 4.) This intensity is recorded on the horizontal diameter of the pole figure n degrees from the circumference in the left half of the pole figure. Similarly, counterclockwise motion of the shaft yields data for the portion of the horizontal diameter in the right half of the pole figure. If the same procedure is followed with the vertical ring moved to a new position in its groove, data are obtained for another diameter of the pole figure, the angular separation from the horizontal diameter being equal to the amount of rotation of the vertical ring. In this way, the entire pole figure may be covered except for the central portion—which is barred when the specimen mount obstructs the x-rays.

The location of data on diameters eliminates the pole figure chart. Only a polar net (Fig. 5) is required, and the same net is, of course, applicable to all materials and all wave-lengths, since the only change in going from one material to another or one wave-length to another is the position of the Geiger counter.

For sheet specimens a correction may be applied according to the formula

$$(I_{\pm\alpha})_{\text{corr.}} = (I_{\pm\alpha})_{\text{obs.}} \times I_0/I_{\pm\alpha},$$

where

$$I_0/I_{\pm\alpha} = \frac{\mu t \exp[-\mu t/\cos\theta]}{\cos\theta} \times \frac{[\cos(\theta\pm\alpha)/\cos(\theta\mp\alpha)]-1}{\exp[-\mu t/\cos(\theta\pm\alpha)]-\exp[-\mu t/\cos(\theta\mp\alpha)]}.$$

This formula takes into account absorption change and change in diffracting volume as the shaft of the sample holder is rotated $\pm\alpha°$ from its zero position. The plus sign refers to counterclockwise and the minus sign to clockwise motion. $I_0/I_{\pm\alpha}$ is the ratio of intensities at 0° and $\pm\alpha°$ for a randomly oriented sample. μ is the coefficient of absorption of the material and t the thickness. Since μ and t only appear as the product μt, a measurement may be made for each sample simply by taking a reading of the Geiger counter using a diffracted beam from any sample whatsoever, then repeating this reading with the sample to be measured placed in front of the Geiger counter window; the μt desired may be calculated from these two readings using the well-known absorption formula

$$I_t = I_0 e^{-\mu t}.$$

(Derivation of the correction formula is given in the appendix.)

To check the validity of this correction formula, data were taken using the AgBr in an x-ray film as a randomly oriented sample. Comparison of observed and corrected intensities is given in Table I. Note that the corrected intensities are

TABLE I.

	AgBr (x-ray film) ⟨220⟩		(10-mil slit at source)		
	1-mm slit at counter (scale 16; time 64)		No slit at counter (scale 32; time 32)		
α	(I_α) obs.	(I_α) corr.	(I_α) obs.	(I_α) corr.	V_α/V_0
−50°	125	209	48	80	1.054
−40	157	211	60	81	0.977
−30	172	209	67	81	0.938
−20	188	213	69	78	0.929
−10	196	210	73	78	0.949
0	208	208	79	79	1.000
10	223	207	84	78	1.092
20	242	204	90	76	1.245
30	246	186	97	74	1.501
40	279	187	102	68	1.965
50	286	169	104	62	2.973

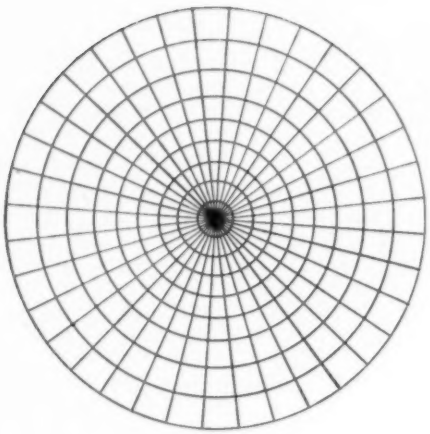


FIG. 5. Stereographic polar net.

equal to the intensity at $\alpha = 0$ for values of α from $-50°$ to $+20°$, after which a decrease appears. This is even true when no slit is used at the counter, so it cannot be caused by the cutting off of part of the diffracted beam by the counter slit. The decrease in corrected intensities parallels a decided increase in diffracting volume V_a (or diffracting path length since the cross section is constant) so it may be due to extinction of the primary beam, or some other function of volume or path length. However, it is unnecessary to use positive values of α in making a pole figure, since the same region may be studied using negative values simply by rotating the vertical ring through 180°. In this way, the unexplained deviations are avoided.

Examples of pole figures determined by the method described above, showing concentration of ⟨100⟩ poles in rolled copper, are given in Figs. 6 to 8. Only one quadrant of the pole figure is reproduced since the other three quadrants may be obtained by mirror reflections of this quadrant. Intensity measurements were taken for every 10° of rotation of the shaft and the vertical ring, and also at positions of intensity maxima when they did not occur at these regular locations. The diameter of the dot at each position is a measure of the intensity relative to the maximum intensity. Ten intensity ranges were used in plotting the figures. The central portion within the dotted line could not be covered because of the obstruction of x-rays by the sample mount. Figure 6 is the ⟨100⟩ pole figure of Cu cold-rolled 99.5 percent, no corrections applied. Figure 7 is the same pole figure after correction according to the formula given above. Figure 8 is the corrected ⟨100⟩ pole figure of 99.5 percent cold-rolled Cu which had been annealed for a half hour at 400°C.

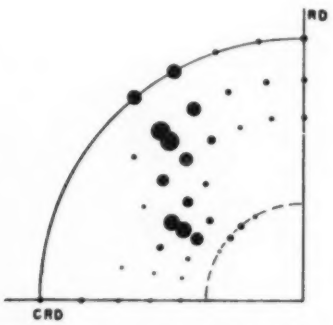


FIG. 6. ⟨100⟩ pole figure of Cu cold-rolled 99.5 percent, no corrections applied.

APPENDIX

Derivation of the Correction Formula

For a sheet specimen the relative intensities of the diffracted beam for different positions of the sample are not true measures of the orientation in the sample since a change in the inclination of sample to beam changes the diffracting volume and the path length within the sample. A correction may be made which eliminates the effects of these changes leaving only crystallite orientation as a cause for change in intensity.

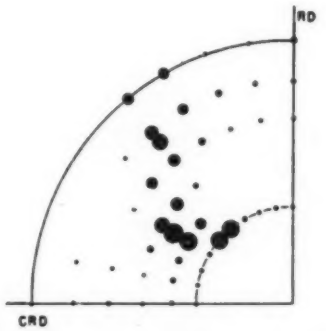


FIG. 7. Corrected ⟨100⟩ pole figure of Cu cold-rolled 99.5 percent.

Assume a sample with random orientation. Rotation of the vertical ring about its axis brings about no change in the inclination of sample to beam, so only rotations of the shaft need be considered. The diffracted intensities for the zero

position of the shaft ($\alpha=0$), and positions attained by counterclockwise ($+\alpha$) and clockwise ($-\alpha$) rotations of the shaft will be calculated.

Figure 9 shows the path of a ray for $\alpha=0$ where A = incident intensity, dI_0 = intensity diffracted from an element of thickness dx, t = thickness of sample, θ = Bragg angle, x = path length within sample before diffraction occurs, $(t/\cos\theta)-x$ = path length of diffracted beam within sample, and $t/\cos\theta$ = total diffracting path length. If μ is

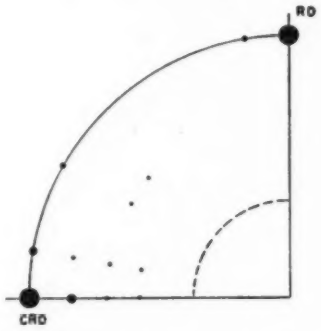


FIG. 8. Corrected ⟨100⟩ pole figure of Cu cold-rolled 99.5 percent and annealed for ½ hr. at 400°C.

the coefficient of absorption of the material, the incident diffracting intensity at distance x is

$$A\exp[-\mu x].$$

Let D be the fraction of incident intensity diffracted at any point in the sample. D is a constant for random orientation. Then, for a diffracting element of thickness dx, the diffracted

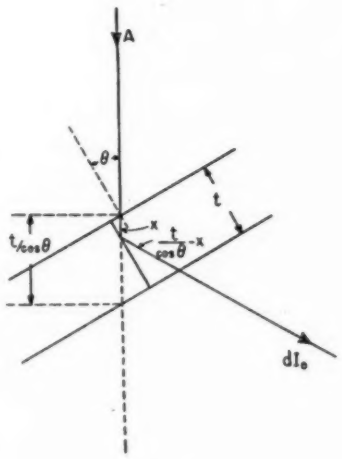


FIG. 9.

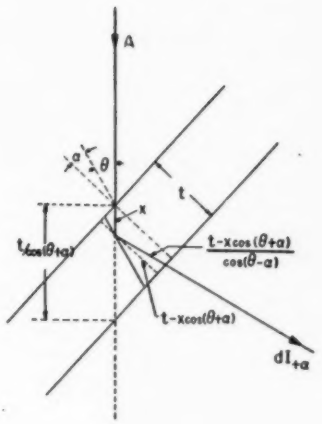


FIG. 10.

intensity transmitted is

$$dI_0=A\exp[-\mu x]Ddx\exp\left[-\mu\left(\frac{t}{\cos\theta}-x\right)\right].$$

To get the total diffracted intensity transmitted, this must be integrated from $x=0$ to $x=t/\cos\theta$:

$$I_0=AD\exp[-\mu t/\cos\theta]\int_0^{t/\cos\theta}dx$$

$$=AD(t/\cos\theta)\exp[-\mu t/\cos\theta].$$

Figures 10 and 11 show the path of a ray for $+\alpha$ and $-\alpha$. The changes in dimensions from the $\alpha=0$ case are shown on the diagrams. A similar treatment gives the result:

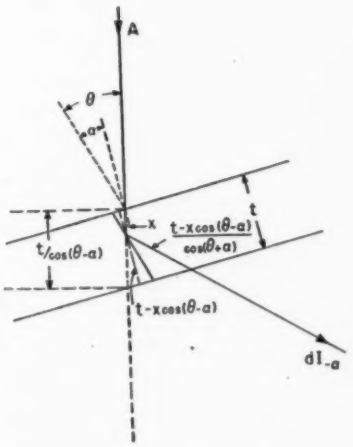


FIG. 11.

$$I_{\pm\alpha}=AD\int_0^{t/\cos(\theta\pm\alpha)}\exp[-\mu x]dx\times\exp\left[-\mu\frac{t-x\cos(\theta\pm\alpha)}{\cos(\theta\mp\alpha)}\right]$$

$$=\frac{AD}{\mu\left[\dfrac{\cos(\theta\pm\alpha)}{\cos(\theta\mp\alpha)}-1\right]}\times[\exp[-\mu t/\cos(\theta\pm\alpha)]-\exp[-\mu t/\cos(\theta\mp\alpha)]],$$

and the factor for changing observed intensity at α to corrected intensity for comparison with observed intensity at $\alpha=0$ is

$$I_0/I_{\pm\alpha}=\frac{\mu t\exp[-\mu t/\cos\theta]}{\cos\theta}\times\frac{[\cos(\theta\pm\alpha)/\cos(\theta\mp\alpha)]-1}{\exp[-\mu t/\cos(\theta\pm\alpha)]-\exp[-\mu t/\cos(\theta\mp\alpha)]}.$$

Effect of Radiation on the Rate of Burning of Solid Fuel Rocket Propellants*

S. S. PENNER**
Allegany Ballistics Laboratory, Cumberland, Maryland
(Received November 21, 1947)

The temperature rise of powder grains in different rocket motors, resulting from the absorption of radiant energy, has been computed by the method of Avery. The calculated results show that more radiant energy is absorbed by the powder grains in large service rockets than in small experimental test motors. The increase in powder temperature before burning, associated with the absorption of radiant energy, leads to an increase in the average rate of burning of the solid fuel charge. The differences in burning rate produced by the absorption of radiant energy amount to less than four percent for the rocket motors commonly employed. Thus burning rate results obtained in small test motors may be applied to predict the approximate behavior of the same propellant in a large service rocket.

INTRODUCTION

THE effect of rocket chamber size on the average burning rate was first discussed by Beek[1] and then by Dresher and McClure.[2] Results similar to those obtained by these authors may be derived in a relatively simple manner by application of Avery's[3] method for the determination of the powder temperature rise associated with the absorption of radiant energy.

Many rockets use solid fuel, perforated, cylindrical powder grains as fuel. A representative grain of this type is shown in Fig. 1. The powder grain is supported by means of a solid wire trap passing through the axial perforation of the cylindrical grain. During burning the entire powder surface is ignited and recedes in a direction normal to itself. There are small differences between the burning rates at different parts of the powder grain. These differences are, among other things, the result of differences in the radiation path length above the burning powder surface. The determination of the dependence of the burning rate on the radiation path length forms the subject of this article.

* This is the *second* of *three articles* abstracted from OSRD report No. 5251, which was written in 1945 at the Allegany Ballistics Laboratory. The work was carried out under OSRD Contract OEMsr-273 with the George Washington University, Washington, D.C. The original paper was reissued as Part IV of OSRD Report No. 5817. A summary of the original paper is given in the Bibliography of Scientific and Industrial Reports, Office of Technical Service, Department of Commerce, **3**, No. 4 (Oct. 1946). The PB number is 34764. The radiation program at the Allegany Ballistics Laboratory was initiated by the Director of Research, Dr. R. E. Gibson. Dr. W. H. Avery, to whom the author is indebted for advice throughout the course of the work, directed the radiation studies. The author is indebted to the Editorial Staff of the Jet Propulsion Laboratory, California Institute of Technology, for material aid in preparing this paper for publication.

** Present address: Jet Propulsion Laboratory, California Institute of Technology, Pasadena, California.

[1] J. Beek, Jr., OSRD Report No. 5817 (Dec. 1945), Part I.

[2] M. J. Dresher and F. T. McClure, OSRD Report No. 5817 (Dec. 1945), Part III.

[3] W. H. Avery, OSRD Report No. 5817 (Dec. 1945), Part II; S. S. Penner, *ibid.*, Part IV; preceding paper: S. S. Penner, J. App. Phys. **19**, 278 (1948).

CALCULATION OF THE TEMPERATURE RISE OF PERFORATED POWDER GRAINS

The following discussion is based on the method of Avery.[3] A cylindrical, perforated, powder grain with outside diameter D_2 and inside diameter D_1 is supported on a trap wire with diameter D_t. The powder grain is contained in a cylindrical powder chamber with inside diameter D_c. The radiant energy incident on the outside surface of the powder grain is designated as I_2, the radiant energy incident on the powder surface from the perforation is I_1 (Fig. 2).

Calculation according to the method of Avery involves the simplifying assumption of greybody emission by the radiating gases. Furthermore, only radiant energy incident in a direction normal to the powder surface is considered.

It is desired to compute the temperature rise, caused by the absorption of radiant energy, in the powder grain at a distance $b \leqslant \frac{1}{2}W$ from the surface of the perforation. Here $W=\frac{1}{2}(D_2-D_1)$ is the web thickness of the powder. The procedure outlined in the previous article leads to the following relations for the radiant energies I_1 and I_2[4]:

$$I_1=I_0[1-\exp\{-k\rho[\tfrac{1}{2}(D_1-D_t)+rt]\}] = I_0\{1-\exp[-\alpha(\beta_1+rt)]\}, \quad (1)$$

$$I_2=I_0[1-\exp\{-k\rho[\tfrac{1}{2}(D_c-D_2)+rt]\}] = I_0\{1-\exp[-\alpha(\beta_2+rt)]\}, \quad (2)$$

where

I_0 = radiant energy emitted by a blackbody radiator at the temperature of the powder gases, in cal./sq. cm/sec.;
k = mass emissivity of the radiating gases, in sq. cm/g, assumed to be independent of wave-length;
ρ = gas density, in g/cc;
r = linear rate of burning of the powder, in cm/sec.;
t = time, in sec.;
$\alpha=k\rho$;
$\beta_1=\frac{1}{2}(D_1-D_t)$;
$\beta_2=\frac{1}{2}(D_c-D_2)$;
D_1, D_2, D_t, D_c, I_1, and I_2 have their previous meaning.

Let K_λ represent the absorption coefficient of the powder in the wave-length region between λ and $\lambda+d\lambda$; $E_{\lambda 1}$ is the radiant energy absorbed in the same wave-length region by the powder strip of width a where a is at any initial distance less than or equal to $\frac{1}{2}W$ from the surface of the perforation; $E_{\lambda 2}$ is the corresponding energy absorbed in a similar thickness a at a distance less than or equal to $\frac{1}{2}W$ from the outside surface of the powder. If K_λ is so large that $e^{-\frac{1}{2}WK_\lambda} \ll 1$ ($K_\lambda \gtrsim 10$ for the propellants used in many rockets), then the products of average burning rate and temperature rise, associated with the absorption of radiant energy, are given by Eqs. (3) and (4), respectively (see Appendix I for details):

$$\bar{r}\Delta T_{\lambda 1}=(2I_{0\lambda}/C\rho_p W) \times\{\tfrac{1}{2}W-\tfrac{1}{2}K_\lambda-[K_\lambda/(K_\lambda-\alpha)] \times[\exp(-\alpha\beta_1)(1-\exp(-\tfrac{1}{2}\alpha W))/\alpha +\exp[-\alpha(\beta_2+\tfrac{1}{2}W)]/(2K_\lambda-\alpha) -\exp[-\alpha\beta_1]/K_\lambda]\}, \quad (3)$$

$$\bar{r}\Delta T_{\lambda 2}=(2I_{0\lambda}/C\rho_p W) \times\{\tfrac{1}{2}W-\tfrac{1}{2}K_\lambda-[K_\lambda/(K_\lambda-\alpha)] \times[\exp(-\alpha\beta_2)(1-\exp(-\tfrac{1}{2}\alpha W))/\alpha +\exp[-\alpha(\beta_1+\tfrac{1}{2}W)]/(2K_\lambda-\alpha) -\exp(-\alpha\beta_2)/K_\lambda]\}. \quad (4)$$

Here $\bar{r}$ is the average burning rate; $I_{0\lambda}$ is the radiant energy emitted in the wave-length region between λ and $\lambda+d\lambda$; $\Delta T_{\lambda 1}$ is the average temperature rise, caused by the absorption of radiant energy in the wave-length region between λ and $\lambda+d\lambda$, for all of the powder grain which was initially at a distance less than or equal to $\frac{1}{2}W$ from the surface of the perforation; $\Delta T_{\lambda 2}$ denotes the corresponding temperature rise for initial distances less than or equal to $\frac{1}{2}W$ from the outside surface of the powder grain; ρ_p is the

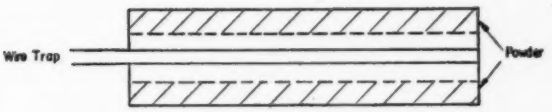


FIG. 1.

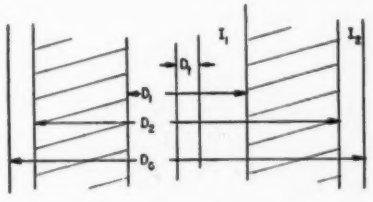


FIG. 2. Cross section of a cylindrical, perforated, powder grain supported on a wire trap.

powder density; C is the specific heat of the solid powder per unit weight.

The values of $\bar{r}\Delta T_1$ and $\bar{r}\Delta T_2$ are obtained by summing the contributions for all wave-length regions, i.e.,

$$\left.\begin{aligned}\bar{r}\Delta T_1&=\textstyle\sum_\lambda \bar{r}\Delta T_{\lambda 1},\\ \bar{r}\Delta T_2&=\textstyle\sum_\lambda \bar{r}\Delta T_{\lambda 2}.\end{aligned}\right\} \tag{5}$$

The average temperature rise, caused by the absorption of radiant energy, is therefore, approximately,

$$\Delta T=\tfrac{1}{2}(\Delta T_1+\Delta T_2), \tag{6}$$

where ΔT_1 and ΔT_2 are obtained from Eqs. (5) by use of Eqs. (3) and (4), respectively.

The determination of ΔT permits the calculation of the effect of radiation path length on the rate of burning of the powder.

APPROXIMATE METHODS FOR THE CALCULATION OF THE EFFECT OF RADIATION ON LINEAR BURNING RATES

The burning rate dependence on radiation path length can be calculated from the known relations between powder temperature and burning rate. A useful approach for numerical evaluation is outlined in the following.

Summation of Eqs. (8) of the preceding report[3] with respect to wave-length leads to Eq. (7):

$$\frac{r_1\Delta T_1 C\rho_p}{I_0}=\sum_\lambda\left[(1-e^{-K_\lambda b})+\frac{K_\lambda e^{-kL_{01}\rho}}{K_\lambda-k\rho}(\exp(-K_\lambda b)-\exp(-kb\rho))\right]. \tag{7}$$

In Eq. (7) r_1 is the burning rate and ΔT_1 is the temperature rise associated with the radiation path length L_{01}. The other symbols have their usual meaning.

At a given pressure Eq. (7) may be written in the form

$$r_1\Delta T_1 C\rho_p/I_0=f(L_{01}). \tag{8}$$

The corresponding relation for a motor with initial path length L_{02} is

$$r_2\Delta T_2 C\rho_p/I_0=f(L_{02}). \tag{9}$$

Division of Eq. (8) by Eq. (9) leads to the result

$$r_1\Delta T_1/r_2\Delta T_2=f(L_{01})/f(L_{02})=\zeta(L_{01},L_{02}), \tag{10}$$

where the function ζ depends on the two path lengths L_{01} and L_{02}.

For many powders the temperature dependence of the burning rate may be expressed by a relation of the form of Eq. (11):[4]

$$r_1=cp^n/[T_i-(T_p+\Delta T_1)]. \tag{11}$$

In Eq. (11) p is the pressure, T_p is the initial powder temperature, and T_i, c, and n are constants characteristic of the powder. Similarly,

$$r_2=cp^n/[T_i-(T_p+\Delta T_2)]. \tag{11a}$$

From Eqs. (11) and (11a) it follows that

$$r_1/r_2=(T_i-T_p-\Delta T_2)/(T_i-T_p-\Delta T_1).$$

Subtracting unity from both sides of the last expression:

$$\begin{aligned}r_1/r_2-1&=(\Delta T_1-\Delta T_2)/(T_i-T_p-\Delta T_1)\\&=(\Delta T_1/\Delta T_2-1)/[(T_i-T_p)/\Delta T_2\\&\qquad-\Delta T_1/\Delta T_2].^{***}\end{aligned} \tag{12}$$

From Eq. (10) it can be seen that

$$\Delta T_1/\Delta T_2=\zeta(L_{01},L_{02})r_2/r_1$$

and, therefore,

$$r_1/r_2=1+(\zeta r_2/r_1-1)/[(T_i-T_p)/\Delta T_2-\zeta r_2/r_1],$$

or

$$r_1/r_2=1+(\zeta-1)\Delta T_2/(T_i-T_p). \tag{13}$$

Let r_2 represent the measured burning rate in an experimental motor with radiation path length L_{02}. Equation (9) may be written as follows:

$$\Delta T_2=f(L_{02})I_0/r_2C\rho_p.$$

Therefore,

$$r_1/r_2=1+(\zeta-1)f(L_{02})I_0/(T_i-T_p)C\rho_p r_2, \tag{14}$$

---

[4] W. H. Avery, R. E. Hunt, and M. N. Donin, OSRD Report No. 5827 (Jan. 1946); W. H. Avery, R. E. Hunt, and L. D. Sachs, OSRD Report No. 5824 (March 1946).

*** This relation is also valid if the burning rate is expressed by a relation of the form (see reference 5)

$$r=(a'+b'p)/(T_i-T_p-\Delta T),$$

where a', b', and T_i are constants characteristic of the powder.

or, more explicitly,

$$\frac{r_1}{r_2}=1+\sum_{\lambda} I_{0\lambda}\left\{\frac{(1-\exp(-K_\lambda b))+\dfrac{K_\lambda \exp(-k\rho L_{o1})}{K_\lambda-k\rho}(\exp(-K_\lambda b)-\exp(-k\rho b))}{(1-\exp(-K_\lambda b))+\dfrac{K_\lambda \exp(-k\rho L_{o2})}{K_\lambda-k\rho}(\exp(-K_\lambda b)-\exp(-k\rho b))}-1\right\}$$

$$\times\frac{(1-\exp(-K_\lambda b))+\dfrac{K_\lambda \exp(-k\rho L_{o2})}{K_\lambda-k\rho}(\exp(-K_\lambda b)-\exp(-k\rho b))}{(T_i-T_p)C\rho_p r_2}.$$

The second term on the right-hand side of the preceding relations may be looked upon as the "radiation correction" to the burning rate resulting from a change in radiation path length. If $L_{o1}=L_{o2}$, then $\zeta=1$ and, therefore, $r_1/r_2=1$. If $L_{o1}>L_{o2}$, then $\zeta>1$ and $r_1>r_2$; if $L_{o1}<L_{o2}$, then $\zeta<1$ and $r_1<r_2$.

The entire radiation correction may be designated as ζ_r, where

$$\zeta_r=(\zeta-1)f(L_{o2})I_0/(T_i-T_p)C\rho_p r_2.$$

Thus Eq. (14) may be written in the convenient form

$$r_1=r_2(1+\zeta_r). \tag{15}$$

COMPARISON OF THEORY AND EXPERIMENT

Consider a perforated powder grain with initial outside path lengths L_{o1} and L_{o2}, respectively. If the absorption coefficient of the powder grain is sufficiently high or if the grain has a sufficiently large web thickness, then the rate of burning of the outside powder surface is affected only by radiant energy incident from the outside, and the rate of burning of the powder surface exposed to the perforation is affected only by radiation from the perforation. However, if a very transparent or else a very thin powder grain is used, then the radiant energy from the outside will contribute to the rate of burning of that part of the powder burning from the surface exposed to the perforation. Similarly, radiant energy from the perforation will contribute to the heating of the outer parts of the cylindrical grain. The radiant energy emitted from the perforation depends only on the dimensions of the powder grain and is therefore independent of motor size. For this reason the difference in outside radiation path length is relatively more effective for opaque or large powder grains. The use of Eq. (14) for a perforated powder grain will therefore result in the determination of an upper limit for the radiation correction if the outside path lengths are used. More nearly correct calculations require the use of Eq. (5). Both methods of calculation are described in the following sections.

TABLE I. Radiation intensity and flame temperature as a function of pressure for T-2 powder.

Pressure, lbs./sq. in.	Gas density, g/cc	Temperature, °K	Radiant energy, cals./sq. cm/sec.
735	0.00649	2170	32.0
1470	0.01280	2180	33.0
2205	0.01910	2175	32.5
2940	0.02524	2175	32.5

Maximum Radiation Correction for Burning Rates in the 115-mm Rocket

The path length L_{o1} in a small experimental test rocket is 1.016 cm. The corresponding value for the 115-mm service rocket is 2.03 cm. For a propellant grain containing 0.02 percent of finely divided carbon, designated as T-2 powder (see Appendix II for the composition of a representative powder), it was found[5] that an absorption

TABLE II. Radiation correction for the burning rate in the 115-mm rocket at various pressures for T-2 powder.

Pressure, lbs./sq. in.	ζ_r for $b=0.10$ cm	ζ_r for $b=0.20$ cm	ζ_r for $b=0.50$ cm
735	0.0342	0.0345	0.0319
1470	0.0325	0.0320	0.0275
2205	0.0247	0.0237	0.0189
2940	0.0184	0.0172	0.0128

[5] B. L. Crawford, Jr. and R. Gledhill in OSRD Report No. 6374 (Dec. 1945); S. S. Penner and F. Daniels in OSRD Report No. 6559, December 1945.

TABLE III. Radiation correction for the burning rate in the 115-mm rocket at various pressures for an undarkened powder grain.

Pressure, lbs./sq. in.	ζ_r at $b=0.10$ cm
735	0.0232
1470	0.0202
2205	0.0154
2940	0.0117

coefficient $K=34$ cm^{-1} held, approximately, for the entire wave-length region from 0.05 to 10.0 microns. The mass emissivity (k) is probably of the order of 40 sq. cm/g.[6] The flame temperature of a given powder may be calculated by standard techniques.[7]

The gas density, theoretical flame temperature, and radiant energy of a blackbody emitter are given in the following Table I for T-2 powder.

The radiation correction to the burning rate (ζ_r) was computed from Eq. (15) by use of the data listed in Table I. Results are summarized in Table II.

Reference to Table II indicates that the maximum radiation correction decreases with increasing pressure. The decrease with pressure is approximately the same for all of the values of b. For any given pressure the value of ζ_r is smallest for the greatest value of b. This is the expected result, since the rate of burning increases during burning so that the powder surface regresses most rapidly at the largest distance from the initial powder surface, thus allowing relatively less time for radiation heating. The data at 735 lbs./sq. in. pressure suggest the existence of an intermediate maximum for ζ_r. The conclusion that the radiation correction to the linear burning rate is of the order of a few percent is in agreement with the views of other authors.[1-3]

Calculations similar to those discussed in the preceding paragraphs have been carried out for an undarkened powder similar in composition to T-2 but not containing added carbon. Here $K=5$ cm^{-1} and $k=40$ sq. cm/g. The initial powder temperature (T_i) was chosen as 12°C,

[6] R. S. Craig, OSRD Report No. 5832 (Nov. 1945).
[7] R. R. Wenner, *Thermochemical Calculations* (McGraw-Hill Book Company, Inc., New York, 1941), pp. 236–242; J. O. Hirschfelder *et al.*, unpublished.

TABLE IV. Ratio of the burning rate in the 115-mm rocket to the burning rate in a small test motor at various pressures for T-2 powder.

Pressure, lbs./sq. in.	r_2/r_1*
735	1.020
1470	1.017
2205	1.011
2940	1.005

* The ratio r_2/r_1 corresponds to an outside radiation path length ratio of 2:1. The inside path length was 0.273 cm in each case.

and the following burning rate expression was used (burning rate expressed in in./sec.):

$r_2=0.726p^{0.683}/(285-T_i)$ for $p>1000$ lbs./sq. in.;

for $p=735$ lbs./sq. in. the burning rate is 0.245 in./sec.[5] The calculated results for $b=0.10$ cm are summarized in Table III.

A comparison of the data shown in Tables II and III, respectively, indicates that the radiation correction to the burning rate is somewhat greater for the more opaque powder. However, the results are of the same order of magnitude. The conclusion may therefore be drawn that the maximum radiation correction is relatively small when the path length is doubled, as is the case in going from a small test motor to the 115-mm rocket.

Radiation Correction for Perforated Charges in the 115-mm Rocket

Equations (5) and (6) should be used to determine the effect of radiation on the burning rate of perforated cylindrical powders. The results of a detailed analysis are summarized in Table IV.

The results shown in Table IV are somewhat lower than those calculated by taking account only of the radiant energy incident from one side. This is in agreement with the previous discussion. The method used for estimating the maximum radiation correction gives values for the ratio of the burning rates at definite distances from the initial powder surface. The more detailed analysis gives average values for the entire powder grain. Nevertheless, the magnitude of the radiation correction as well as its change with pressure are found to be similar.

The preceding considerations indicate that burning rates determined in small test motors

will be only slightly lower than burning rates in larger service rockets, provided all of the factors which affect the burning rate of powders, other than radiation, are held constant.

APPENDIX I. AVERAGE TEMPERATURE RISE OF PERFORATED POWDER GRAINS

The radiant energy absorbed by a cylindrical shell of width a in the inner half of the perforated powder grain in the wave-length region between λ and $\lambda+d\lambda$ is $E_{\lambda 1}$ per unit area. From Eqs. (1) and (2) it is easy to show that $E_{\lambda 1}$ is given by the following relation:

$$E_{\lambda 1}=I_0(1-\exp(-K_\lambda a))\times\left\{\int_0^t[1-\exp(-\alpha(\beta_1+rt))]\times\exp(-K_\lambda X_1)dt+\int_0^t[1-\exp(-\alpha(\beta_2+rt))]\times\exp(-K_\lambda X_2)dt\right\}. \quad \text{(A-1)}$$

Here X_1 represents the distance of the inside burning surface from the powder layer which was initially at a distance $b<\frac{1}{2}W$ from the inner powder surface; X_2 is the corresponding distance of the outside burning surface from the same layer. The following boundary values must be satisfied:

$$\begin{aligned} &X_1=b-rt \text{ for } t>0,\\ &X_1=0 \text{ when } t=b/r,\\ &X_1=b \text{ when } t=0, \end{aligned} \quad \text{(A-2)}$$

and

$$\begin{aligned} &X_2=W-b-rt \text{ for } t>0,\\ &X_2=\tfrac{1}{2}(D_2-D_1)-b=W-b \text{ when } t=0,\\ &X_2=W-2b \text{ when } t=b/r. \end{aligned} \quad \text{(A-3)}$$

The burning rates of the powder from the inside and from the outside are assumed to be the same for approximate calculations.

Combining Eqs. (A-1), (A-2), and (A-3) leads to the relation

$$E_{\lambda 1}=I_0(1-\exp(-K_\lambda a))\times\left\{\int_0^{b/r}[1-\exp(-\alpha(\beta_1+rt))]\times\exp(-K_\lambda(b-rt))dt+\int_0^{b/r}[1-\exp(-\alpha(\beta_2+rt))]\times\exp(-K_\lambda(W-b-rt))dt\right\}. \quad \text{(A-4)}$$

Integrating Eq. (A-4) and finding the temperature rise in the usual manner[3] leads to the result

$$\frac{\Delta T_{\lambda 1}C\rho_p r}{I_0}=[\exp(K_\lambda b)-1]\times[\exp(-K_\lambda b)+\exp(-K_\lambda b')]-\frac{K_\lambda}{K_\lambda-\alpha}[\exp[(K_\lambda-\alpha)b]-1]\times[\exp(-K_\lambda b-\alpha\beta_1)+\exp(-K_\lambda b'-\alpha\beta_2)], \quad \text{(A-5)}$$

where $b'=W-b$. Application of the preceding considerations to a cylindrical shell of width a in the outer part of the powder grain gives the relation

$$\frac{\Delta T_{\lambda 2}C\rho_p r}{I_0}=[\exp(K_\lambda c)-1]\times[\exp(-K_\lambda c)+\exp(-K_\lambda c')]-\frac{K_\lambda}{K_\lambda-\alpha}[\exp(K_\lambda-\alpha)c-1]\times[\exp(-K_\lambda c-\alpha\beta_2)+\exp(-K_\lambda c'-\alpha\beta_1)], \quad \text{(A-6)}$$

where c represents the distance of the cylindrical shell under consideration from the outside powder surface and $c'=W-c$. Equations (A-5) and (A-6) are identical except for the interchanging of β_1 and β_2 and the replacement of b by c and b' by c', respectively.

A distance weighted average of the products $r\Delta T_{\lambda 1}$ and $r\Delta T_{\lambda 2}$ can be found from Eqs. (A-5) and (A-6) by integrating Eq. (A-5) from 0 to $\frac{1}{2}W$ with respect to b and multiplying the result by $2/W$ and by integrating Eq. (A-6) from 0 to $\frac{1}{2}W$ with respect to c and multiplying the result again by $2/W$. Equations (A-7) and (A-8) are found in this manner:

$$\bar{r}\Delta T_{\lambda 1}=\frac{2I_0}{C\rho_p W}\left\{\tfrac{1}{2}W+\frac{1}{K_\lambda}(\exp(-\tfrac{1}{2}K_\lambda W)-1)+\frac{\exp(-K_\lambda W)}{2K_\lambda}\times(\exp(K_\lambda W)-2\exp(\tfrac{1}{2}K_\lambda W)+1)-\frac{K_\lambda}{K_\lambda-\alpha}\left[-\frac{\exp(-\alpha\beta_1)}{\alpha}\times(\exp(-\tfrac{1}{2}\alpha W)-1)+\frac{\exp(-K_\lambda W-\alpha\beta_2)}{2K_\lambda-\alpha}\times(\exp[(2K_\lambda-\alpha)\tfrac{1}{2}W]-1)+\frac{\exp(-\alpha\beta_1)}{K_\lambda}(\exp(-\tfrac{1}{2}K_\lambda W)-1)-\frac{1}{K_\lambda}\exp(-\alpha\beta_2-K_\lambda W)\times(\exp(\tfrac{1}{2}K_\lambda W)-1)\right]\right\}, \quad \text{(A-7)}$$

and

$$\bar{r}\Delta T_{\lambda 2}=\frac{2I_0}{C\rho_p W}\left\{\tfrac{1}{2}W+\frac{1}{K_\lambda}(\exp(-\tfrac{1}{2}K_\lambda W)-1)+\frac{\exp(-K_\lambda W)}{2K_\lambda}\times(\exp(K_\lambda W)-2\exp(\tfrac{1}{2}K_\lambda W)+1)-\frac{K_\lambda}{K_\lambda-\alpha}\left[-\frac{\exp(-\alpha\beta_2)}{\alpha}\times(\exp(-\tfrac{1}{2}\alpha W)-1)+\frac{\exp(-K_\lambda W-\alpha\beta_1)}{2K_\lambda-\alpha}\times(\exp((2K_\lambda-\alpha)\tfrac{1}{2}W)-1)+\frac{\exp(-\alpha\beta_2)}{K_\lambda}(\exp(-\tfrac{1}{2}K_\lambda W)-1)-\frac{1}{K_\lambda}\exp(-\alpha\beta_1-K_\lambda W)\times(\exp(\tfrac{1}{2}K_\lambda W)-1)\right]\right\}. \quad \text{(A-8)}$$

If K_λ is so large that

$$\exp(-\tfrac{1}{2}K_\lambda W)\ll 1,$$

then Eqs. (A-7) and (A-8) reduce to Eqs. (3) and (4), respectively, for the numerical values of α, β_1, β_2, k, and W likely to be encountered in solid fuel rocket motors.

APPENDIX II. NOMINAL COMPOSITION OF JPT POWDER

Nitrocellulose (13.25 percent N)	58.6 percent
Nitroglycerin	38.9 percent
Ethyl centralite†	1.0 percent
Potassium sulfate	1.3 percent
Diphenylamine	0.2 percent
Volatiles (added)	1.0 percent

† Ethyl centralite is diethyl-diphenyl-urea.

Observations on Arc Discharges at Low Pressures

M. J. KOFOID*
Westinghouse Research Laboratories, East Pittsburgh, Pennsylvania
(Received November 13, 1947)

It has been observed in these laboratories (J. Slepian and E. J. Haverstick, Phys. Rev. **33**, 52 (1929); J. Slepian, W. E. Berkey, and M. J. Kofoid, J. App. Phys. **13**, 113 (1942)) that in a certain range of pressures, of the order of millimeters, short arcs burn with low apparent cathode current densities. The melting of the electrodes is then greatly reduced.

In this paper are described spectrographic tests to determine the presence of electrode material in the arc. It was found that as the gas pressure was increased, for an arc of given current and length, at a critical pressure the amount of cathode metal in the arc, as indicated by spectral line density, suddenly increased. Presumably, boiling of the cathode metal took place at this pressure. This critical pressure was found to increase as the arc length was reduced.

In this range where melting of the cathode occurred, it was found that the molten spot was at the center of the cathode for magnetic cathode material but at the edge of the cathode for non-magnetic material.

INTRODUCTION

SHORT arcs in gas at reduced pressure have shown promise for use in practical discharge gaps. The interest in these arcs of 1 mm to about 2 cm in length, burning in the common gases at pressures in the range of a fraction of one mm to about 15 cm of mercury, lies in the fact that they may have low apparent cathode spot current densities. By apparent current density is meant the total current divided by the arced-over area. The measured current densities of 60-cycle, 1000-ampere r.m.s. arcs of one-half cycle duration burning in air at a few cm pressure on copper cathodes were less than 1000 amperes per sq. cm.[1] In air at atmospheric pressure, the cathode current density is of the order of several thousand amperes per sq. cm, or more.

A result of the low current density is extremely small electrode burning effects with currents of several thousand amperes for one-half cycle duration.[1] A second important characteristic of these arcs is that in low pressure air a short arc space recovers dielectric strength at least as fast as the same length of arc in air at atmospheric pressure. However, the maximum dielectric strength is much greater at atmospheric pressure.

This paper describes basic studies made to provide results useful in the practical design of low gas-pressure arcing devices.**

QUALITATIVE SPECTROGRAPHIC STUDY OF PRESENCE OF CATHODE MATERIAL IN ARC DISCHARGE

If a half-cycle, 1600-crest-ampere arc is drawn between two plane electrodes of the same metal separated by a few mm in gas at low pressure, no visible surface disturbance except oxidation will take place, provided the gas density is in a certain low range. If the short spacing is maintained but the gas density is gradually increased, a condition will be reached where signs of electrode melting due to the arc will be found on the surface of the cathode. The gas density may be increased considerably further before the anode surface will show signs of being affected. In fact, the copper cathodes showed bright cathode arced-over areas and the spectrographic record indicated the presence of large quantities of copper in the arc, while the copper anodes have been observed to become heavily coated with what appeared to be oxide.

Spectrograms were taken, therefore, to determine if there were sudden changes in the amount of metal boiled off the cathode as the gas density was increased. Such transformations

* Now at Department of Electrical Engineering, Oregon State College, Corvallis, Oregon.

[1] J. Slepian, W. E. Berkey, and M. J. Kofoid, J. App. Phys. **13**, 113 (1942).

** The spectrographic study of the amount of cathode metal in the arc was carried out by the writer at Stanford University as a doctorate thesis. Incidental observations during this work suggested the remainder of the work, which was carried out at the Westinghouse Laboratories.

were found to take place with definite noticeable suddenness. Then the spectrographic investigation was extended to determine if the transformation could be correlated with parameters of pressure, electrode separation, or current for different gases and electrode materials.

Description of the Experimental Tests

With particular electrode spacings and particular values of current, the gas pressure was varied to determine the value below which the amount of metal in the arc was markedly decreased.

In the tests the range of pressures was from a fraction of a millimeter to 200 mm of mercury and the range of spacings was from 1.0 mm to 18 mm. The tests were made in air.

Previous work with arcs of 1000-ampere crest indicated arc cathode current densities to be less than 1000 crest-amperes per sq. cm for half-cycle arcs between large plane copper electrodes separated 2 mm in air at a pressure of approximately 1 cm. The arcs were a.c. of one-half cycle duration. Most such arcs were found to have a lower current density of 500 to 600 amperes per sq. cm. In the spectrographic study, peak current densities of more than 350, 550, and 700 crest-amperes per sq. cm were obtained by electrode area limitation with the selected test currents of 1000, 1600, and 2000 crest-amperes, respectively. The 60-cycle current waves were nearly sinusoidal.

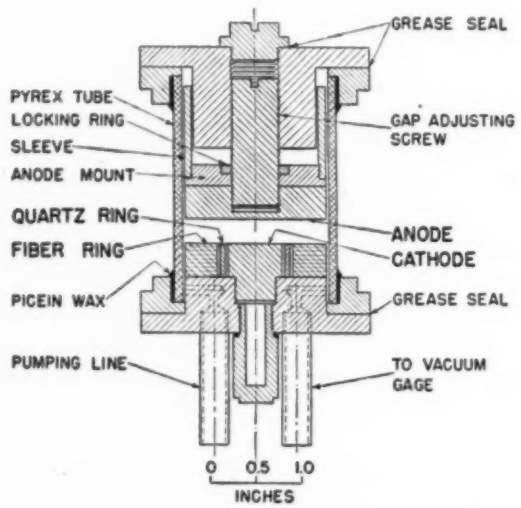


FIG. 1. Test chamber for investigating properties of low gas-pressure arc discharge gaps.

The arc voltage was not measured under each condition of test. Extensive oscillographic recording in previous work, involving tests under similar conditions on all cathode materials used in the present work, showed uniform arc voltages of 30 volts, or less.

The arcing chamber built for these investigations was designed to confine the arc, but yet to allow visual and photographic observations of the arcs to be made. The arcing chamber is shown in the drawing of Fig. 1.

The phenomena in an arc would be changed if confining walls were near the periphery of the arc. The construction of this arcing chamber minimized this type of disturbance for short arcs. The arcs were prevented from nearing the glass container walls by using a cathode of much smaller diameter than the inside diameter of the tube. A quartz ring surrounded the cathode.

Gas density is an important parameter in arc phenomena and should be maintained as constant as possible throughout the duration of the arc. For this reason the walls of the anode and cathode structures were fitted closely with the wall of the glass tube; the passages were of capillary dimensions. The amount of gas that could escape from the main gap region in the short time of 0.01 second, due to the sudden pressure increase caused by the arc, was, therefore, small.

Copper anodes of a single size were used. The cathodes had a diameter of 0.47 that of the anode. The various cathodes used differed in material only.

The electrical circuit was arranged to break

TABLE I. Spectrographic data for short low air-density arcs.

Cathode metal	Gap (mm)	Current (crest-amp.)	Pressure range (mm)	Appearance of cathode metal spectral lines in arc: Pressure range (mm)	Appearance of cathode metal spectral lines in arc: Nature of appearance
Iron	1	1600	1–200	100–120	Start quite abruptly
	2	1000	1–30		No lines appear
	2	1600	0.1–60	20–25	Start abruptly
	2	2000	1–30	20–30	Abrupt change
	5	1600	1–30	4–6	Small change
	10	1000	1–40	4–6	Small change
	10	1600	1–20	8–10	Abrupt change
	18	1600	1–12	10–12	Abrupt change
Copper	1	1600	1–40		No lines appear
	2	1600	1–40	30–40	Abrupt change
	5	1600	1–40	20–30	Small change
	10	1600	1–30	15–20	Abrupt change
	18	1600	0.1–12	8–10	Abrupt change
Tungsten	1	1600	10–100	40–50	Abrupt change

down the gap by a high voltage condenser discharge, and the 110-volt r.m.s., 60-cycle voltage produced a power-follow current. The arc normally did not restrike after being extinguished at the first current zero. The arc current of every operation during spectrographic recording was observed visually on an oscilloscope for its maximum value.

The spectrographic records were obtained by means of a simple prism spectrograph. The slit was parallel to the axis of the arcing chamber. Light from the entire length of the arc was permitted to fall upon the slit in all cases. The records were made on Eastman Super XX film. Each spectrogram was obtained by exposing the film to the light from 10 successive arc discharges.

Experimental Results

A tabulation of the arc investigated and of the results obtained from the spectrograms is given in Table I. The spectrograph was incapable of resolving individual lines, and only unresolved groups of lines could be identified in the spectrograms. By "appearance" of groups of spectral lines is here meant either of two things. There may have been no cathode metal line groups distinctly visible in the negative until they suddenly appeared with appreciable density. On the other hand, the line groups may have been very faintly visible at lower pressures but suddenly increased considerably in density. In either case it was the point of sudden increase that was sought, and this has been termed the "appearance" of the lines.

The sudden strong increase in intensity is interpreted as indicating boiling, not melting. At pressures of a little below the pressures at which the changes in line group intensities appeared, melting was found on electrodes.

As an illustration of how the data of Table I were obtained, the spectrograms of the 1600-ampere arc with an iron cathode and a 10-mm gap are presented as an example. The spectrograms are shown in Fig. 2. (The printed photograph indicates the results only in a very general way by making evident the general nature of the spectrogram and showing the major changes in the spectral line groups; small changes in line group density were visible in the negative.) Between the pressures of 8 and 10 mm, groups of iron lines have appeared strongly at wavelengths of about 5200, 4700, 4450, and 4270A. In the negative these line groups are first apparent at 4-mm pressure, but remain quite faint until the pressure is increased to 10 mm where a pronounced density change occurs. The wavelength scale agrees with the bottom spectrogram only. The other spectrograms are displaced sideways by small random amounts caused in moving the film; the relative position of the right-hand edge of the oscillogram is indicative of the shifts that have occurred.

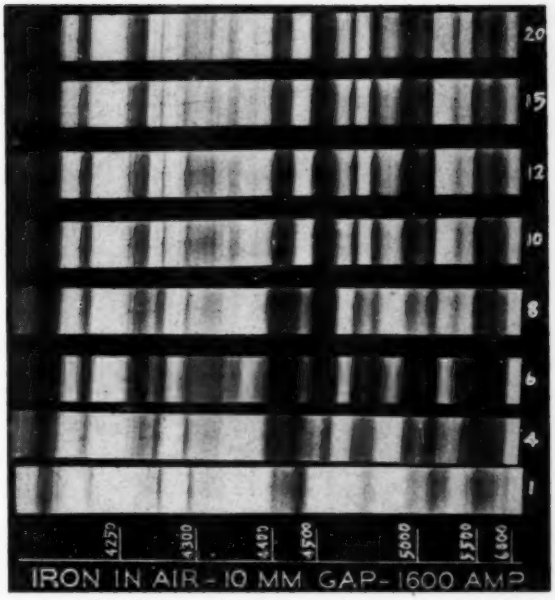


FIG. 2. Typical spectrogram showing change in cathode metal line group densities when boiling of cathode metal occurs. In this spectrogram the change occurred at a pressure between 8 and 10 mm. Pressures are indicated by the numbers at the right. The lines in the print are the reverse of the lines in the negative.

With 1600-ampere arcs in 1.0-mm gaps, no cathode metal lines were strong enough to be seen in the negative for pressures up to 40–50 mm for iron, copper, or tungsten. Tungsten line groups appeared abruptly at 40–50 mm. Iron line groups appeared at 100–120 mm. Copper lines would be expected to appear at about the same pressure as iron lines since in other tests copper closely followed iron.

The abrupt appearance of strong cathode line groups occurs at much lower pressures when the gap length is increased. This is clearly shown by

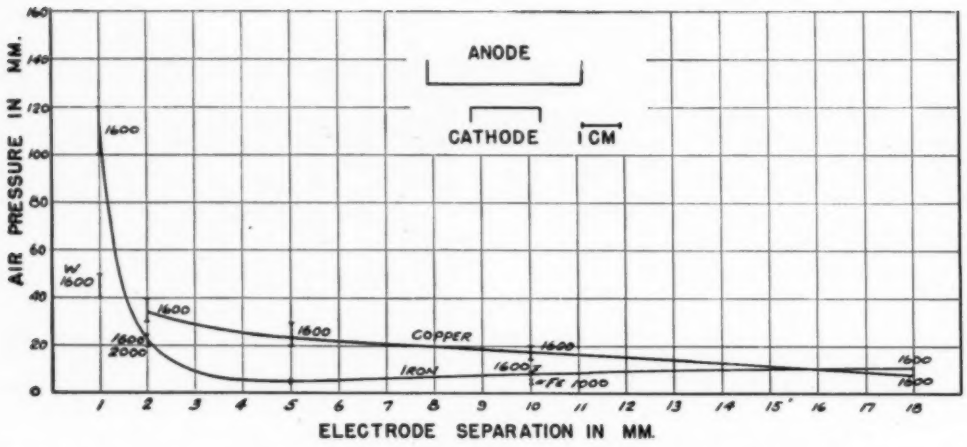


FIG. 3. Arc length *vs.* pressure at which spectral lines changed in intensity suddenly. Arcs burning in air. Total crest current is indicated beside each point.

the curves of Fig. 3 plotted from the data of Table I.

The cathode metals tested, arranged in the order of first appearance of strong spectral line groups, with increasing pressure are Al, W, Fe, and Cu; that is, Al lines appeared in arcs at a pressure lower than that at which W lines first appeared. Any general difference between Cu and Fe seems questionable. Strong Al line groups were present in 1600-crest-ampere arcs at as low as 0.1-mm pressure for a 1.0-mm gap.

The cathode metals tested, arranged in the order of first appearance of appreciable surface melting with increase of pressure, are Al, W, Cu, and Fe; that is, Al melted at a pressure far below the pressure at which W melted.

A formula for the temperature rise of an electrode surface is given by Slepian.[2] If all the heat of an arc is assumed to go into the cathode, the temperature rise T in degrees C of the cathode surface in t seconds is:

$$T=\frac{2W(t)^{\frac{1}{2}}}{4.18(kc\delta)^{\frac{1}{2}}},$$

where W is the watts per sq. cm input, k is the heat conductivity in calories per cm³ per degree C, c is the heat capacity in calories per cu. cm per degree C, and δ is the density of the cathode material in grams per cu. cm. Slepian's formula indicates that the time required for the surface of the cathode to attain a particular temperature is proportional to the quantity $T(kc\delta)^{\frac{1}{2}}$. The metals tested, arranged in descending order of their values of the product $T(kc\delta)^{\frac{1}{2}}$, with the temperature rise T taken as the rise from room temperature to the melting point of the cathode metal, are W, Cu, Fe, and Al. That is, with a constant rate of heat input, the time required to heat the cathode surface to its melting point is

FIG. 4. Effect on cathode surfaces of a half-cycle 1600-crest-ampere arc in air. Pressure (p) and separation (d) are both in mm. Boiling did not occur in the tests of the electrodes shown in photographs "a" and "b."

[2] J. Slepian, J. A. I. E. E. **48**, 92 (1929).

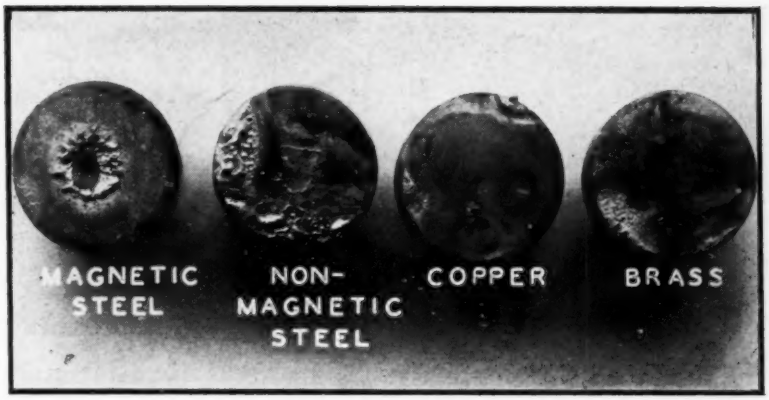


FIG. 5. Cathodes after repeated 1300-crest-ampere arcs of one-half cycle duration. Arcs drawn in hydrogen at a pressure of 5 cm. Electrode separation 1.0 cm. Cathodes were of solid metal 1 cm thick.

greater for W than for Cu. The order, W, Cu, and Fe is the same as given above for the first appearance of appreciable surface melting with increase of pressure,

In Figs. 4a and 4b are shown electrodes which have been cathodes for a single half-cycle 1600-ampere arc. The aluminum cathode was deeply melted under the low pressure and short-gap conditions of $p=1$ mm and $d=1.0$ mm. Under these same conditions the 1600-ampere arc is seen to have merely lightly oxidized the surfaces of the iron and copper cathodes. The copper and iron cathodes after the ($p=2$ mm) and ($d=10$ mm) operation, shown in Fig. 4c, illustrate the type of melting characteristic of copper and of iron with low pressure arcs when the pressure is about the lowest at which melting was obtained in air with $d=10$ mm. Copper forms a melted crater which can be seen surrounded by a discolored oxidized area. (Tungsten melts in a similar manner but the crater is much more shallow.) Iron does not melt deeply but may form beads of metal on the surface, as occurred on the electrode of Fig. 4c.

POSITIONING OF CATHODE SPOTS ON DIFFERENT METALS

The action of a low current density arc on a cathode of magnetic material differs from the action on non-magnetic materials with respect to the preferred location of the cathode spot. Electrodes were arced under conditions of separation and gas pressure which permitted some melting of the cathode surface. The nature of the results is indicated by the test electrodes shown in Fig. 5 and Fig. 6, which are discussed later.

Description of the Experimental Tests

The current in all tests discussed was about 1300 crest-amperes. The spacing between electrodes was 1.0 cm, and hydrogen gas was used at a pressure of 5 cm of mercury. (Similar results were observed in tests made in air.)

All the tests were made in the test chamber of Fig. 1, with special precautions taken to have the current leads approach the top and bottom of the chamber approximately on the center line so that the magnetic field of the current in the leads would produce a minimum action on the arc.

In the laboratory in which these tests were carried out it was necessary to use a higher voltage of 220 volts, 60 cycle in order to obtain the desired currents. Although arcs of half-cycle duration were desired, numerous reignitions occurred. Therefore, the markings and melting on the test electrodes are presented to show only the manner in which the cathode spot tends to concentrate or spread out. The relative amounts of surface disturbance in this group of tests are not intended to be indicative or comparative.

Experimental Results

The photograph of Fig. 5 shows that the entire area of the surface of the solid copper cathode has been arced over, but that the melting was always confined to the periphery. In contrast, the photograph of the solid steel electrode shows that its entire area has been arced over, but that melting is confined to a circular central area.

With cathodes of non-magnetic steel or of brass, the behavior is very similar to that of copper. (The non-magnetic steel was about 80 percent iron and had a permeability of 5 at 100

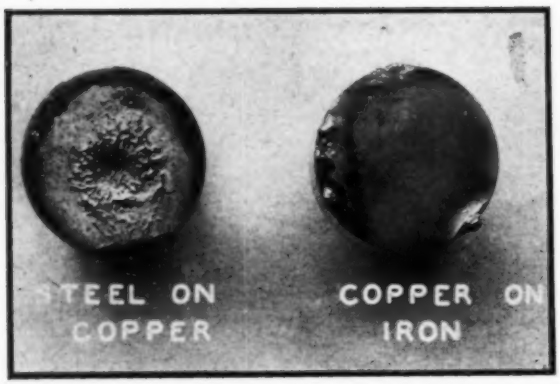


FIG. 6. Cathodes after repeated 1300-crest-ampere arcs of one-half cycle duration. Arcs drawn in hydrogen at a pressure of 5 cm. Electrode separation 1.0 cm. Surface metal 0.8 mm thick silver-soldered to body metal 9 mm thick.

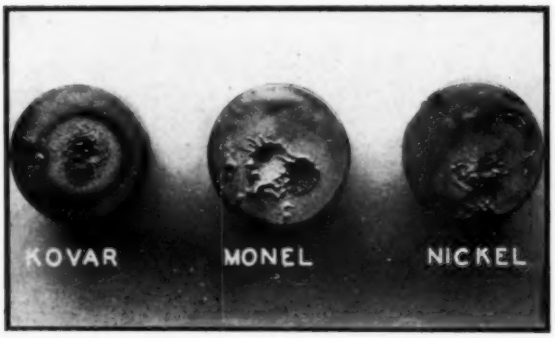


FIG. 7. Cathodes after repeated 1300-crest-ampere arcs normally of one-half cycle duration. (Reignitions occurred in the tests on Kovar and nickel.) Arcs drawn in hydrogen at 5-cm pressure. Electrode separation 1 cm. The Kovar, monel, and nickel electrodes were 2.4, 2.4, and 0.8 mm thick, respectively, and were soldered to thick brass bodies.

oersteds.) Tests were also made with cathodes constructed with a $\frac{1}{32}$-inch thick surface layer of one metal silver-soldered to a much thicker body of a second metal. The photographs of Fig. 6 show that the performance with the laminated electrodes of steel on copper and of copper on steel was practically identical with that of solid electrodes of the same metal.

The results of tests on Kovar, monel, and nickel are shown in the photographs of Fig. 7. Kovar has a permeability approaching that of iron. Monel and nickel both have much lower permeabilities. The permeabilities at 100 oersteds for Kovar, nickel, and monel are roughly 170, 60, and 40, respectively. (The permeability of monel virtually disappears at about 100 degrees C.) With Kovar the arc was located at the center of the cathode except when the arc reignited. In the tests on monel no reignition occurred; the arc is seen to be located at other positions in addition to locating at the center. The results on nickel do not indicate any preferred location of the cathode arc terminal.

CONCLUSIONS

1. 1300-crest-ampere arcs of half-cycle duration in air or hydrogen, under conditions of pressure $p=50$ mm and spacing $d=10$ mm, tend to concentrate symmetrically in the center of a circular iron cathode but always tend to move out toward the periphery if the cathode metal is copper or brass. Brass and non-magnetic iron cathodes give results similar to those obtained with copper. The greater the permeability, the more nearly similar are the results to those obtained with iron.

2. 1300-crest-ampere arcs of one-half cycle duration in air under conditions of $p=5$ mm and $d=2$ mm produce serious melting with a copper cathode, slight melting with an iron cathode, but a smooth matt surface with a brass cathode. That is, brass is the most resistant to surface melting and puddling by low pressure arcs.

3. For each spacing, with a given gas, there appears to be a pressure below which the amount of cathode metal in the arc decreases markedly.

The Stresses in a Plate Containing an Overlapped Circular Hole

CHIH-BING LING
California Institute of Technology, Pasadena, California
(Received November 17, 1947)

This paper deals with the stresses in a plate containing a hole whose rim is composed of two equal circular arcs. The stress function concerned is constructed by means of an even integral solution of biharmonic equation expressed in bipolar coordinates. The parameters involved in the solution are determined from the given boundary conditions with the aid of Fourier transforms. Three fundamental stress systems are discussed. Expressions of the stress along the rim of hole are derived and, in particular, values of the maximum stress are calculated. The results are then plotted in conjunction with those of a plate containing two equal circular holes, which are obtained by the present author in a previous paper. It may be interesting to compare the the methods used in both solutions.

INTRODUCTION

IN a previous paper,[1] the author has investigated the stresses in a plate of infinite size, containing a group of two equal circular holes, under the action of several fundamental stress systems in the plane of the plate. The maximum stresses in the plate, which always occur at the rims of holes, are evaluated against the distance between the holes. The limiting case in which the holes are tangential to each other is also investigated.

The previous solution, however, does not cover the case in which the distance between the holes further decreases beyond the preceding limiting case. The holes then overlap each other to form a single hole whose rim is composed of two equal circular arcs. In this paper a solution of the plate containing such an overlapped hole will be given in terms of bipolar coordinates. The same fundamental stress systems will be considered to act in the plane of the plate.

It is noted that the two arcs of the overlapped hole as described above are equal and each is greater than a semicircle. Nevertheless, the present solution applies equally well to cases in which the arcs are each less than a semicircle, i.e., the hole is in a lenticular shape.

EVEN INTEGRAL SOLUTION OF BIHARMONIC EQUATION

In order to render the present results directly comparable with those in the previous paper, the bipolar coordinates (ξ, η) will be defined alternatively by the equation

$$x+iy=-a\cot\tfrac{1}{2}i(\xi+i\eta), \tag{1}$$

such that the two poles of the coordinates are now located on the y axis at the points $(0, \pm a)$, and

$$x=J\sin\eta, \quad y=J\sinh\xi, \tag{2}$$

where

$$a/J=\cosh\xi-\cos\eta. \tag{3}$$

It may be mentioned that the coordinate η is multi-valued. However, in our problem η will be restricted to vary between $-\pi$ and π with the line $\eta=\pi$ connecting the poles as the branch line.

The biharmonic equation $\nabla^4\chi=0$ which must be satisfied by the stress function χ then transforms to

$$\left(\frac{\partial^4}{\partial\xi^4}+2\frac{\partial^4}{\partial\xi^2\partial\eta^2}+\frac{\partial^4}{\partial\eta^4}-2\frac{\partial^2}{\partial\xi^2}+2\frac{\partial^2}{\partial\eta^2}+1\right)\frac{\chi}{J}=0. \tag{4}$$

It can be readily shown that an integral solution of this equation, which is even in both ξ and η, may be obtained in the following form:[2]

$$\chi_1/J=\int_0^\infty F_n(\eta)\cos n\xi\, dn, \tag{5}$$

where

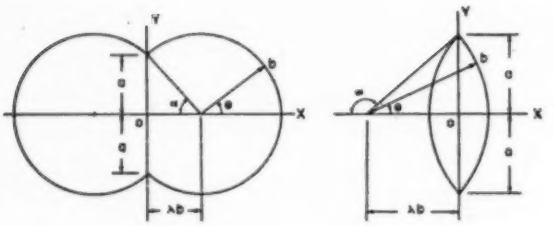
$$F_n(\eta)=f_n\sin\eta\sinh n\eta+k_n\cos\eta\cosh n\eta, \tag{6}$$

f_n and k_n being parametric functions of n.

Furthermore, when a whole plate containing

[1] C. B. Ling, "On the stresses in a plate containing two circular holes," J. App. Phys. 19, 77 (1948).

[2] C. B. Ling, "On the stresses in a notched plate under tension," J. Math. and Phys. 26, No. 4 (1947).



(1) (2)

FIG. 1. (1) $\alpha<\pi/2$, $(\lambda>0)$. (2) $\alpha>\pi/2$, $(\lambda<0)$.

an overlapped hole, as the present problem, is dealt with, it appears convenient to include also in the solution (5) an even biharmonic function:

$$\chi_2/J=K(\cosh\xi-\cos\eta)\log\frac{\cosh\xi-\cos\eta}{\cosh\xi+\cos\eta}, \quad (7)$$

where K is a parametric coefficient. It is readily shown that

$$\frac{\cosh\xi+\cos\eta}{\cosh\xi-\cos\eta}=\frac{x^2+y^2}{a^2}. \quad (8)$$

Hence, the function χ_2/J represents in fact a center of radial pressure at the origin $(x, y)=0$.

THE METHOD OF SOLUTION

Let us now consider the plate in question. The two arcs which form the rim of hole will be represented by $\eta=\pm\alpha$. They intersect at the poles of the coordinates and their centers lie symmetrically on the x axis (see Fig. 1).

If the hole were absent, the stress system in the plate will be supposed to be specified by an even stress function χ_0. The method of solution when the hole is present is to add to χ_0 the even biharmonic functions χ_1 and χ_2, as shown in (5) and (7), together with a factor aT. The factor is introduced to render the parameters involved in the functions dimensionless. That is to say that the stress function χ is constructed in the form:

$$\chi=\chi_0+aT(\chi_1+\chi_2). \quad (9)$$

The parameters involved are then adjusted to satisfy the boundary conditions of the plate, which are as follows: First, the stresses derived from $aT\chi_1$ and $aT\chi_2$ all vanish at infinity $(\xi, \eta)=0$, such that at infinity the stresses derived from χ_0 are undisturbed by the presence of the hole. Second, there is no traction acting on the rim of hole, i.e., that the stresses $\widehat{\eta\eta}$ and $\widehat{\xi\eta}$ derived from χ both vanish identically when $\eta=\pm\alpha$.

It is well known that the stresses derived from a stress function χ in terms of bipolar coordinates are given by

$$\left.\begin{aligned}
a\widehat{\xi\xi}&=\left\{(\cosh\xi-\cos\eta)\frac{\partial^2}{\partial\eta^2}-\sinh\xi\frac{\partial}{\partial\xi}\right.\\
&\qquad\left.-\sin\eta\frac{\partial}{\partial\eta}+\cosh\xi\right\}\frac{\chi}{J},\\
a\widehat{\eta\eta}&=\left\{(\cosh\xi-\cos\eta)\frac{\partial^2}{\partial\xi^2}-\sinh\xi\frac{\partial}{\partial\xi}\right.\\
&\qquad\left.-\sin\eta\frac{\partial}{\partial\eta}+\cos\eta\right\}\frac{\chi}{J},\\
a\widehat{\xi\eta}&=-(\cosh\xi-\cos\eta)\frac{\partial^2}{\partial\xi\partial\eta}\left(\frac{\chi}{J}\right).
\end{aligned}\right\} \quad (10)$$

The first boundary condition thus requires that $(\chi_1+\chi_2)/J=0$ when $(\xi, \eta)=0$. Since $\chi_2/J=0$ when $(\xi, \eta)=0$, this condition leads to

$$\int_0^\infty k_n dn=0. \quad (11)$$

The second boundary condition requires that the stress function χ must satisfy the following relations at $\eta=\pm\alpha$.

$$\{(\cosh\xi-\cos\alpha)(\partial^2/\partial\xi^2)-\sinh\xi(\partial/\partial\xi)+\cos\alpha\}\times[\chi/J]_\alpha-\sin\alpha[(\partial/\partial\eta)(\chi/J)]_\alpha=0. \quad (12)$$

$$-(\cosh\xi-\cos\alpha)(\partial/\partial\xi)[(\partial/\partial\eta)(\chi/J)]_\alpha=0. \quad (13)$$

Equation (13) yields at once

$$(\partial/\partial\xi)[(\partial/\partial\eta)(\chi/aTJ)]_\alpha=0. \quad (14)$$

By differentiating with respect to ξ and with the aid of the above relation, Eq. (12) then gives, apart from a factor,

$$\{(\partial^3/\partial\xi^3)-(\partial/\partial\xi)\}[\chi/aTJ]_\alpha=0. \quad (15)$$

Equations (11), (14), and (15) provide sufficient number of equations for the determination of f_n, k_n, and K, provided that the basic stress function χ_0 is given.

EVALUATION OF PARAMETERS

As in the previous paper, three fundamental stress systems will be considered to act in the plane of the plate. The basic stress function χ_0 in each case is as follows: (1) when the plate is under an all-around tension T,

$$\chi_0=\tfrac{1}{2}T(x^2+y^2)$$

or

$$\chi_0/aTJ=\tfrac{1}{2}(\cosh\xi+\cos\eta), \tag{16}$$

(2) when under a longitudinal tension T,

$$\chi_0=\tfrac{1}{2}Ty^2$$

or

$$\chi_0/aTJ=\tfrac{1}{2}\sinh^2\xi/(\cosh\xi-\cos\eta), \tag{17}$$

and (3) when under a transverse tension T,

$$\chi_0=\tfrac{1}{2}Tx^2$$

or

$$\chi_0/aTJ=\tfrac{1}{2}\sin^2\eta/(\cosh\xi-\cos\eta). \tag{18}$$

In order to keep χ/J finite throughout the plate, any multiple of $(\cosh\xi-\cos\eta)$ will be removed from χ/J. This is permissible since the removal does not affect the stresses. The function $\cos\eta$ alone then gives an all-around tension in the plate. Thus, χ_0/aTJ may be written together as follows:

$$\chi_0/aTJ=\delta\cos\eta\pm\tfrac{1}{2}\sin^2\eta/(\cosh\xi-\cos\eta), \tag{19}$$

where $\delta=0$ in the transverse tension case, or $\delta=1$ in the other two cases. In regard to ambiguous signs, the upper one goes with the longitudinal tension case and the lower one with the transverse tension case. In the all-around tension case the terms with ambiguous signs are all absent. For the sake of brevity, these conventions will be used henceforth to distinguish the three cases.

Now, by (9), Eq. (14) yields

$$\int_0^\infty nF_n'(\alpha)\sin n\xi\,dn = \frac{K\sin 2\alpha\sinh 2\xi}{(\cosh\xi-\cos\alpha)(\cosh\xi+\cos\alpha)^2} \pm\frac{\sin\alpha\sinh\xi(\cos\alpha\cosh\xi-1)}{(\cosh\xi-\cos\alpha)^3}. \tag{20}$$

The prime sign has the usual meaning of differentiation. It appears that the function $\cos\eta$ contributes nothing to the right-hand side of Eq. (20).

By integrating with respect to ξ, Eq. (15) becomes

$$\{(\partial^2/\partial\xi^2)-1\}[\chi/aTJ]_\alpha=\text{const.}=C. \tag{21}$$

Consequently, by (9) and using (19) for χ_0/aTJ, this equation leads to

$$\int_0^\infty (n^2+1)F_n(\alpha)\cos n\xi\,dn = -C-(\delta-2K)\cos\alpha+K\cos\alpha\log\frac{\cosh\xi-\cos\alpha}{\cosh\xi+\cos\alpha}+\frac{2K\cos^2\alpha}{\cosh\xi+\cos\alpha} -\frac{K\sin^2 2\alpha}{(\cosh\xi-\cos\alpha)(\cosh\xi+\cos\alpha)^2}\mp\frac{\sin^2\alpha}{2}\left\{\frac{3\cos\alpha}{(\cosh\xi-\cos\alpha)^2}-\frac{2\sin^2\alpha}{(\cosh\xi-\cos\alpha)^3}\right\}. \tag{22}$$

The right-hand side of the equation vanishes when $\xi=\infty$, provided that

$$C=-(\delta-2K)\cos\alpha. \tag{23}$$

To facilitate subsequent analysis, C will be so adjusted. This is justified only if a function $(\delta-2K)\cos\eta$ for an all-around tension is removed from χ/aTJ. The removal implies that the new $[\chi_1/J]_\alpha$ will vanish when $\xi=\infty$. This leaves Eq. (20) unchanged, but the condition (11) must now be replaced by

$$\int_0^\infty k_n\,dn=\delta-2K. \tag{24}$$

By applying Fourier's sine and cosine transforms to Eqs. (20) and (22), respectively, we then have

$$nF_n'(\alpha)=\frac{2K\sin 2\alpha}{\pi}\int_0^\infty \frac{\sinh 2\xi \sin n\xi d\xi}{(\cosh\xi-\cos\alpha)(\cosh\xi+\cos\alpha)^2}$$

$$\pm\frac{2\sin\alpha}{\pi}\int_0^\infty \frac{(\cos\alpha\cosh\xi-1)\sinh\xi\sin n\xi d\xi}{(\cosh\xi-\cos\alpha)^3}, \quad (25)$$

$$(n^2+1)F_n(\alpha)=\frac{2K\cos\alpha}{\pi}\int_0^\infty\left(\log\frac{\cosh\xi-\cos\alpha}{\cosh\xi+\cos\alpha}\right)\cos n\xi d\xi+\frac{4K\cos^2\alpha}{\pi}\int_0^\infty\frac{\cos n\xi d\xi}{\cosh\xi+\cos\alpha}$$

$$-\frac{2K\sin^2 2\alpha}{\pi}\int_0^\infty\frac{\cos n\xi d\xi}{(\cosh\xi-\cos\alpha)(\cosh\xi+\cos\alpha)^2}$$

$$\mp\frac{\sin^2\alpha}{\pi}\int_0^\infty\left\{\frac{3\cos\alpha}{(\cosh\xi-\cos\alpha)^2}-\frac{2\sin^2\alpha}{(\cosh\xi-\cos\alpha)^3}\right\}\cos n\xi d\xi. \quad (26)$$

By means of Cauchy's theorem of residues, it is not difficult to show that

$$\int_0^\infty\frac{n\cos n\xi d\xi}{\cosh\xi-\cos\alpha}=\int_0^\infty\frac{\sinh\xi\sin n\xi d\xi}{(\cosh\xi-\cos\alpha)^2}=\frac{n\pi\sinh n(\pi-\alpha)}{\sinh n\pi\sin\alpha}, \quad (27)$$

$$\int_0^\infty\frac{\sinh\xi\sin n\xi d\xi}{\cosh^2\xi-\cos^2\alpha}=\frac{\pi\sinh n((\pi/2)-\alpha)}{2\cosh(n\pi/2)\cos\alpha}. \quad (28)$$

By differentiating both sides with respect to α, (27) yields successively

$$\int_0^\infty\frac{n\cos n\xi d\xi}{(\cosh\xi-\cos\alpha)^2}=2\int_0^\infty\frac{\sinh\xi\sin n\xi d\xi}{(\cosh\xi-\cos\alpha)^3}=n\pi\frac{n\cosh n(\pi-\alpha)\sin\alpha+\sinh n(\pi-\alpha)\cos\alpha}{\sinh n\pi\sin^3\alpha}, \quad (29)$$

$$\int_0^\infty\frac{\cos n\xi d\xi}{(\cosh\xi-\cos\alpha)^3}=\pi\frac{(n^2+1)\sinh n(\pi-\alpha)\sin^2\alpha+3n\cosh n(\pi-\alpha)\sin\alpha\cos\alpha+3\sinh n(\pi-\alpha)\cos^2\alpha}{2\sinh n\pi\sin^5\alpha}. \quad (30)$$

Also, by replacing α with $\pi-\alpha$, (27) and (29) give, respectively,

$$\int_0^\infty\frac{n\cos n\xi d\xi}{\cosh\xi+\cos\alpha}=\int_0^\infty\frac{\sinh\xi\sin n\xi d\xi}{(\cosh\xi+\cos\alpha)^2}=\frac{n\pi\sinh n\alpha}{\sinh n\pi\sin\alpha}, \quad (31)$$

$$\int_0^\infty\frac{\cos n\xi d\xi}{(\cosh\xi+\cos\alpha)^2}=\frac{\pi(n\cosh n\alpha\sin\alpha-\sinh n\alpha\cos\alpha)}{\sinh n\pi\sin^3\alpha}. \quad (32)$$

Again, from (27) and (29), integration with respect to α yields

$$\int_0^\infty\left(\log\frac{\cosh\xi-\cos\alpha}{\cosh\xi+\cos\alpha}\right)\cos n\xi d\xi=-\frac{\pi\sinh n((\pi/2)-\alpha)}{n\cosh(n\pi/2)}. \quad (33)$$

Here the constant of integration is zero.

With these integrals, we thus deduce

$$nF_n'(\alpha)\sinh n\pi = 4K\{n\sinh n\alpha\cos\alpha+\sinh(n\pi/2)\sinh n((\pi/2)-\alpha)\sin\alpha\}$$
$$\mp n\{n\cosh n(\pi-\alpha)\sin\alpha-\sinh n(\pi-\alpha)\cos\alpha\}, \quad (34)$$

$$n(n^2+1)F_n(\alpha)\sinh n\pi = 4K\{n^2\cosh n\alpha\cos\alpha-n\cosh(n\pi/2)\sinh n((\pi/2)-\alpha)\sin\alpha$$
$$-\sinh(n\pi/2)\sinh n((\pi/2)-\alpha)\cos\alpha\}\pm n(n^2+1)\sinh n(\pi-\alpha)\sin\alpha. \quad (35)$$

Consequently, f_n and k_n are solved in terms of K as follows:

$$f_n=\frac{4K}{n^2+1}\left(\frac{1}{2}\coth\frac{n\pi}{2}-\frac{\sinh^2 n\alpha+\sin^2\alpha}{\sinh 2n\alpha+n\sin 2\alpha}\right)\mp\left(\coth n\pi-\frac{2\cosh^2 na}{\sinh 2n\alpha+n\sin 2\alpha}\right), \quad (36)$$

$$k_n=-\frac{4K}{n(n^2+1)}\left(\frac{1}{2}\coth\frac{n\pi}{2}-\frac{n^2+1}{\sinh n\pi}-\frac{\sinh^2 n\alpha-n^2\sin^2\alpha}{\sinh 2n\alpha+n\sin 2\alpha}\right)\pm\frac{2n\sin^2\alpha}{\sinh 2n\alpha+n\sin 2\alpha}. \quad (37)$$

Equation (24) supplies the necessary condition for the determination of K. By using the following integral from Cauchy's theorem of residues,

$$\int_0^\infty\left(\frac{1}{2}\coth\frac{n\pi}{2}-\frac{n^2+1}{\sinh n\pi}\right)\frac{dn}{n(n^2+1)}=\frac{1}{4}\int_{-\infty}^{\infty}\frac{(\cosh n\pi-2n^2-1)dn}{n(n^2+1)\sinh n\pi}=\sum_{n=1}^{\infty}\frac{1}{4n^2-1}=\frac{1}{2}, \quad (38)$$

we thus find

$$4K\int_0^\infty\frac{(\sinh^2 n\alpha-n^2\sin^2\alpha)dn}{n(n^2+1)(\sinh 2n\alpha+n\sin 2\alpha)}=\delta\mp 2\sin^2\alpha\int_0^\infty\frac{ndn}{\sinh 2n\alpha+n\sin 2\alpha}. \quad (39)$$

It may be noted that the integrals in (27) to (33) can be reversed by means of Fourier's transforms. For instance, the first integral in (27) may be reversed thus:

$$\frac{\sin\alpha}{\cosh\xi-\cos\alpha}=2\int_0^\infty\frac{\sinh n(\pi-\alpha)\cos n\xi dn}{\sinh n\pi}. \quad (40)$$

With the aid of some such integrals it can be shown that

$$2\cos\eta+(\cosh\xi-\cos\eta)\log\frac{\cosh\xi-\cos\eta}{\cosh\xi+\cos\eta}$$
$$=2\int_0^\infty\left(\frac{n\sin\eta}{\sinh(n\pi/2)}+\frac{\cos\eta}{\cosh(n\pi/2)}\right)\frac{\sinh n((\pi/2)-\eta)\cos n\xi dn}{n(n^2+1)}-4\cos\eta\int_0^\infty\frac{n\cosh n\eta\cos n\xi dn}{(n^2+1)\sinh n\pi}. \quad (41)$$

Consequently, remembering that a function $(\delta-2K)\cos\eta$ is removed from χ/aTJ, we reach the following result without difficulty.

$$\frac{\chi}{aTJ}=4K\int_0^\infty\frac{n\cosh n(\alpha-\eta)\sin\eta+\sinh n(\alpha-\eta)\cos\eta}{n(n^2+1)(\sinh 2n\alpha+n\sin 2\alpha)}\sinh n\alpha\cos n\xi dn$$
$$-4K\sin\alpha\int_0^\infty\frac{n\cosh n\eta\sin(\alpha-\eta)+\sinh n\eta\cos(\alpha-\eta)}{(n^2+1)(\sinh 2n\alpha+n\sin 2\alpha)}\cos n\xi dn$$
$$\pm 2\int_0^\infty\frac{n\cosh n\eta\sin(\alpha-\eta)\sin\alpha-\cosh n\alpha\sinh n(\alpha-\eta)\sin\eta}{\sinh 2n\alpha+n\sin 2\alpha}\cos n\xi dn. \quad (42)$$

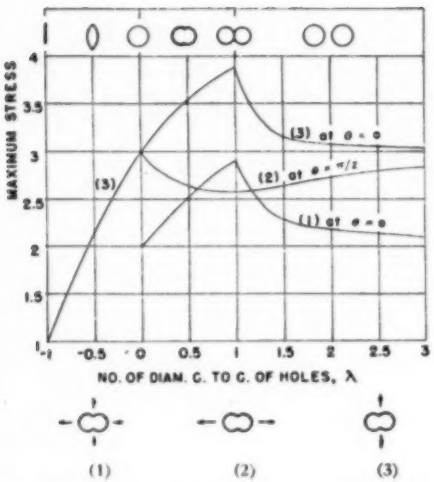


FIG. 2. Maximum stress for $T=1$. (1) All-around tension. (2) Longitudinal tension. (3) Transverse tension.

STRESS ALONG RIM OF HOLE

To find the stress along the rim of hole, we have, by subtracting the first two equations in (10) and by virtue of $[\eta\hat{\eta}]_\alpha=0$,

$$\frac{1}{T}[\xi\hat{\xi}]_\alpha=(\cosh\xi-\cos\alpha)\left[\left(\frac{\partial^2}{\partial\eta^2}-\frac{\partial^2}{\partial\xi^2}+1\right)\frac{\chi}{aTJ}\right]_\alpha$$

$$=4(\cosh\xi-\cos\alpha)\sin\alpha\times\int_0^\infty\frac{2K\mp n(n-\cot\alpha\coth n\alpha)}{\sinh 2n\alpha+n\sin 2\alpha}\times\sinh n\alpha\cos n\xi\,dn, \quad (43)$$

where K is given in (39).

For convenience, the points on the rim of hole will also be specified by λ and θ, as shown in Fig. 1. They are connected to the bipolar coordinates (ξ, α) by

$$\cos\alpha=\lambda,\quad \cosh\xi=(1+\lambda\cos\theta)/(\lambda+\cos\theta). \quad (44)$$

It is readily seen that the maximum stress in the plate occurs on the rim of hole at the following points:

Stress system	For $\alpha<\pi/2$ $(\lambda>0)$	For $\alpha>\pi/2$ $(\lambda<0)$
(1) All-around tension	maximum at $\xi=0$ $(\theta=0)$	infinite at $\xi=\infty$ $(\theta=\pi-\alpha)$
(2) Longitudinal tension	maximum at $\theta=\pi/2$ nearly	infinite at $\xi=\infty$
(3) Transverse tension	maximum at $\xi=0$	maximum at $\xi=0$

In the longitudinal tension case when $\alpha<\pi/2$, the actual point of maximum stress is at θ slightly less than $\pi/2$. For the sake of simplicity, however, the stress at $\theta=\pi/2$ may well be taken as the maximum stress without any appreciable error. The two instances of "infinite" stress are anticipated theoretically, although they may not be so detrimental in practice.

THE LIMITING CASE $\alpha=0$: TWO TANGENTIAL HOLES

The limiting case $\alpha=\pi$ corresponds to a slit in the plate. The case $\alpha=\pi/2$ corresponds to a single circular hole. Both results are well known. On the other hand, the limiting case $\alpha=0$ corresponds to two complete circular holes tangential to each other. In this case, Eq. (39) becomes in the limit,

$$4(\alpha^2K)\int_0^\infty\frac{(\sinh^2\phi-\phi^2)d\phi}{\phi^3(\sinh 2\phi+2\phi)}=\delta\mp 2\int_0^\infty\frac{\phi d\phi}{\sinh 2\phi+2\phi}. \quad (45)$$

The maximum stresses corresponding to the three stress systems are equal to the following limits:

$$\frac{1}{T}\hat{\xi\xi}_0=4(\alpha^2K)\int_0^\infty\frac{\sinh\phi d\phi}{\sinh 2\phi+2\phi},$$

$$\frac{1}{T}\hat{\xi\xi}_0=8(\alpha^2K)\int_0^\infty\frac{\sinh\phi\cos\phi d\phi}{\sinh 2\phi+2\phi}+4\int_0^\infty\frac{\phi(\cosh\phi-\phi\sinh\phi)\cos\phi d\phi}{\sinh 2\phi+2\phi},$$

$$\frac{1}{T}\hat{\xi\xi}_0=4(\alpha^2K)\int_0^\infty\frac{\sinh\phi d\phi}{\sinh 2\phi+2\phi}+2\int_0^\infty\frac{\phi(\phi\sinh\phi-\cosh\phi)d\phi}{\sinh 2\phi+2\phi}. \quad (46)$$

TABLE I. Maximum stress for $T=1$.

λ	α	All-around tension (at $\theta=0$)	Longitudinal tension (at $\theta=\pi/2$)	Transverse tension (at $\theta=0$)	Remark
1	0°	2.894	2.569	3.869	Two tangential holes
0.5	60°	2.524	2.630	3.493	
0	90°	2.000	3.000	3.000	A single hole
−0.5	120°	*	*	2.101	
−1	180°	*	*	1.000	A slit

They are identical with the results obtained in the previous paper as a limiting case from a different consideration.

NUMERICAL RESULTS

Values of the maximum stress are calculated by evaluating the integrals in (43) and (39). They may be integrated numerically by means of Gregory formula,[3] or otherwise. The results are shown in Table I.

In the table, the cases marked with the asterisk (*) are governed by the stress concentration at the re-entrant corners of the lenticular hole, where $\theta = \pm(\pi - \alpha)$. Figure 2 shows the results graphically. By combining with the previous results of two equal circular holes, it is seen that the curves can now be extended beyond $\lambda = 1$.

It may be mentioned that the transverse tension case was discussed by Weinel[4] a few years ago, but his paper was not available in this country until recently. His treatment, however, is slightly different. In the limiting case of two tangential holes he gave the maximum stress as 3.84, whereas the present result is 3.869.

[3] E. T. Whittaker and G. Robinson, *Calculus of Observations* (Blackie and Son, Limited, London, 1937), p. 143.

[4] E. Weinel, "Die Spannungserhöhung durch Kreisbogenberben," Zeits. angew. Math. Mech. **21**, 228–230 (1941).

Minimum Detectable Absorption in Microwave Spectroscopy and an Analysis of the Stark Modulation Method

W. D. HERSHBERGER
Radio Corporation of America Laboratories, Princeton, New Jersey
(Received November 18, 1947)

It is shown that the minimum detectable absorption coefficient for an isolated spectral line as measured in a matched wave guide of optimum length is

$$[4kTN\Delta f/(P_0/e)]\alpha_c,$$

where N is the noise factor and Δf the band width of the receiver, α_c is the attenuation constant for the wave guide, and P_0/e is the average microwave power available at the receiver. The equivalence between cavity and wave guide methods is developed. The Stark modulation method of Hughes and Wilson is analyzed, assuming a quadratic Stark effect, and expressions are developed for the modulation factor of a microwave at twice and four times the frequency of the Stark field. For a matched guide the modulation factor is $\alpha_0 l/4$. Also the shift in line position is calculated, and finally, the directions in which improvement in sensitivity may be expected are indicated.

RECENT work in microwave spectroscopy has resulted in an extension of the list of absorbing gases and in improvements in technique both in measuring the frequency of absorption lines and in the detection of weak lines. In the attempt to measure small absorption coefficients, several lines of attack have been followed. Some workers have used long-matched wave guides but there is an upper limit to the length of guide that yields an increase in sensitivity with increase in length. Bleaney and Penrose[1] have made good use of small gas samples by enclosing them in a high Q resonant cavity operating in a low order mode, but here the upper limit to sensitivity is set by the realizable Q of such cavities which is of the order of 10^4.

Becker and Autler[2] and Lamb,[3] proceeding along somewhat similar lines, realize a Q of the order of 10^6 by the use of a cubical resonant cavity 8 feet on an edge. In studies on line shape or location, it is convenient to work at pressures of 1 mm to 0.01 mm of mercury and a cavity with a volume of 1.6×10^7 cm^3 is not well adapted for this sort of work. Moreover, this large cavity works in a region of high mode density and ex-

[1] B. Bleaney and R. P. Penrose, Proc. Roy. Soc. **A189**, 358 (1947).

[2] G. E. Becker and S. H. Autler, Phys. Rev. **70**, 300 (1946).

[3] W. E. Lamb, Jr., Phys. Rev. **70**, 308 (1946).

hibits many resonances in the band of interest. To surmount this difficulty, the Columbia group used both mode mixing and also sampled the field inside their resonantor at 360 points. Hughes and Wilson[4] have employed a Stark modulation method in a relatively short section of wave guide with excellent results. In this method the position of the line itself is varied cyclically at a low radio-frequency rate. This artifice permits to a large degree the experimental segregation of effects which are due to system frequency selectivity from those which are due to the selective absorption of the gas. The method lends itself well to the detection of small absorption lines but as one attempts to measure frequency with precision-difficulty is experienced, first, because a perturbing field is used altering line position which must be adequately taken into account, and second, because the frequency selectivity of the gas chamber distorts the line so its location is in doubt. A superficially similar system has been in use at these laboratories and was described by the writer at the Ohio State University symposium on molecular spectroscopy on June 12, 1947. A similar system has been used and described by Gordy and Kessler.[5] In this method the klystron frequency is swept 20 Mc at a 60-cycle rate and simultaneously is swept perhaps 40 kc at a 100-kc rate. A 100-kc receiver follows the microwave detector. The method permits fair discrimination between gas selectivity and system selectivity, due primarily to the narrowness of the absorption lines at reduced pressure. In the parlance of radio engineering, the absorption line is used as a discriminator. In addition, a fairly large, high order mode, resonant cavity method has been developed in which Stark modulation is used and which has given good results in measuring small absorption coefficients. The cavity makes effective use of a gas sample larger than was used in the Hughes-Wilson method, but of course the sample is much smaller than that used by the Columbia group.

The problem of detecting and measuring an absorption coefficient involves, in essence, noting the change in power fed through a microwave transmission path containing the gas as we vary a critical parameter. In a simple wave guide or cavity system, gas is admitted to, then evacuated from, the transmission path and the parameter varied is pressure; in the popular sweeping technique,[6,7] microwave frequency is varied and change in transmission as a function of frequency is observed, while in the Stark modulation method the position of the absorption line is varied cyclically by application of an electric field. The purpose of the present paper is to discuss the factors which limit measurements of small absorption coefficients in any method and, further, to present an analysis of the Stark modulation method with the view of showing how it may be used to obtain data for calculating both the magnitude of the absorption coefficient and the correction to line position needed in this type of experiment.

MINIMUM DETECTABLE ABSORPTION

The factors which limit the detection and measurement of small absorption lines find their simplest interpretation in the transmission of power through a matched wave guide. Let the physical length of such a guide be L and its equivalent length be l, where the relationship between these two lengths is determined by the mode in use. α_g is the power absorption coefficient for plane waves for the gas in the guide, while α_c is a corresponding absorption coefficient used to account for losses in the guide walls or in other dielectric material in the guide. The change in power ΔP observed in a receiver is

$$\Delta P = P_0 e^{-\alpha_c l}[1 - e^{-\alpha_g l}] \tag{1}$$

if we experimentally vary the appropriate parameter, as for example, pressure, to obtain the maximum possible power change. When $\alpha_g l \ll 1$, (1) may be written

$$\Delta P = P_0 e^{-\alpha_c l} \alpha_g l. \tag{2}$$

$P_0 e^{-\alpha_c l}$ represents power available for actuating a receiver. ΔP, which is to be measured, attains a maximum value for a critical length of wave guide, namely,

$$l = \alpha_c^{-1}. \tag{3}$$

In the copper wave guide most commonly used at 1.25 cm, the critical length is approximately 01

[4] R. H. Hughes and E. B. Wilson, Jr., Phys. Rev. **71**, 562 (1947).
[5] W. Gordy and M. Kessler, Phys. Rev. **72**, 644 (1947).
[6] W. E. Good, Phys. Rev. **69**, 539 (1946).
[7] C. H. Townes, Phys. Rev. **70**, 665 (1946).

meters and, if we add an electrode and long insulating strips as in the Hughes-Wilson equipment, the optimum length is thereby decreased. The workers using guide sections longer than optimum, in order to realize large fractional changes in received power due to gas absorption, have to accept the penalties which go with a low power level at the receiver. The importance of maximizing ΔP rather than any one of its factors becomes apparent when it is recalled that for satisfactory detection, ΔP must be equal to or greater than $4kT\Delta fN(P)$ where Δf is the band width of the receiver and $N(P)$ its noise factor. There are instances when it is desirable to use guide lengths greater or less than optimum. Thus, if the noise arising in a microwave generator is the limiting factor rather than receiver noise, or if the receiver overloads or limits with the power available through length α_c^{-1} from a given generator, a guide length greater than optimum is indicated. Again, if the noise factor $N(P)$ is reduced in a given receiver by increasing the power above P_0/e, a length shorter than α_c^{-1} will give best results. When the critical guide length is employed, the minimum absorption coefficient one may detect is

$$\alpha_{\min}=[4kTN\Delta f/(P_0/e)]\alpha_c. \tag{4}$$

RESONANT CAVITY

Two types of cavity are in use: in the first, the high Q essential for measuring small absorption coefficients is realized by the choice of a shape giving a favorable volume to surface ratio; in the second, the need for an electrode for application of an electric field dictates a structure which is conveniently treated as a section of wave guide, working in a high order mode, and terminated to give rise to pronounced reflections at its two ends. In this second type, the Q decreases with increasing length instead of tending to a constant value as might be expected from considerations based on the ratio of stored to dissipated energy. The theory of this type of resonator is accordingly developed with the view of enabling one to estimate the optimum size for a gas chamber and to interpret experimental results.

Figure 1 depicts a section of wave guide in which E_1 represents an electric field component in a wave traveling from left to right inside the guide, E_2 represents the corresponding component of an oppositely traveling wave, α represents the total power absorption coefficient for the wave in the guide, and β is the propagation function for the mode considered. We write

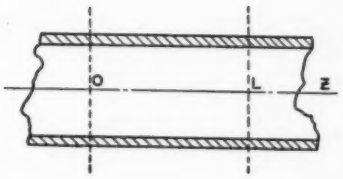


FIG. 1. Wave guide section.

$$E_1=\frac{E_P}{1-e^{-\alpha l}}e^{-(\alpha z/2)}e^{j(\omega t-\beta z)}, \tag{5a}$$

$$E_2=\frac{-E_P}{1-e^{-\alpha l}}e^{-(\alpha/2)(2l-z)}e^{j(\omega t-2\beta l+\beta z)}. \tag{5b}$$

The amplitudes and relative phase of these waves have been adjusted so that when $Z=L$, $E_1+E_2=0$ and when $z=0$,

$$E_1+E_2=\frac{E_Pe^{j\omega t}}{1-e^{-\alpha l}}[1-e^{-\alpha l}e^{-j2\beta L}]. \tag{6}$$

When $L=q\lambda_g/2$, where q is an integer, the quantity in the bracket in (6) is real and

$$E_1+E_2=E_pe^{j\omega t},$$

which represents the driving field. For this case, the guide section acts as a resonant cavity. Let us define Δ as a length such that when

$$L=(q\lambda_g/2)\pm\Delta,$$

the quantity in the bracket in Eq. (6) has an angle $\pm\pi/4$. We now define the Q of the cavity

$$Q\equiv L/2\Delta. \tag{7}$$

If $2\Delta\ll L$, it may be shown that

$$Q=\beta L/e^{\alpha L}-1. \tag{8}$$

For relatively short guide sections such that

$$\alpha L\ll 1,$$

we obtain the customary expression used in resonator work, namely,

$$Q=2\pi/\alpha\lambda, \tag{9}$$

which is independent of L. However, for very

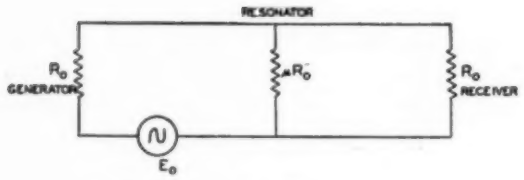


FIG. 2. Equivalent circuit of resonator.

long guide sections (8) reduces to

$$Q=\beta L/e^{\alpha L}, \tag{10}$$

and the denominator in this expression becomes, of course, the controlling factor as L becomes large. α in Eqs. (5) to (10) represents the total power absorption coefficient, that is, $\alpha=\alpha_c+\alpha_g$. If we take Q_c as characteristic of the cavity with no gas in it, and Q_g as arising from the gas alone, and if $Q_g \gg Q_c$, the resulting Q due to all losses is given by

$$Q=Q_c[1-(Q_c/Q_g)] \tag{11a}$$

for resonators whose Q is given by Eq. (9). On the other hand, for resonators whose Q is given by Eq. (10), and if $\alpha_g L \ll 1$,

$$Q=Q_c[1-\alpha_g L]. \tag{11b}$$

The quantity of importance is the change in power received through a resonator whose Q is altered by introduction of an absorbing gas. In the system of interest a generator which matches a wave guide is used in conjunction with a receiver which also matches the guide, but these two components are coupled to each other through the resonant cavity by similar but separate coupling holes. It is now shown that the change in received power which ensues on introduction of an absorbing gas attains a maximum value when the magnitude of the coupling is such that the initial Q of the cavity is halved.

Figure 2 shows the equivalent circuit used in this analysis. R_0 is the guide impedance, μR_0 is the impedance of the resonator referred to the guide employing an equivalent turns ratio r. μR_0 thus consists of two impedances in parallel R_c/r^2 and R_g/r^2 where R_c is the shunt impedance of the resonator and R_g the equivalent impedance arising from the gas. Hence, $\mu=v/[1+(\alpha_g/\alpha_c)]$ where $v=R_c/r^2R_0$. The power dissipation in the receiver is

$$\frac{E_0^2}{R_0}\left(\frac{u}{2u+1}\right)^2=P_L.$$

For no resonator losses, the load dissipation attains a maximum value:

$$P_0=E_0^2/4R_0.$$

It follows that

$$P_L=4P_0\left(\frac{v}{2v+1}\right)^2\left[1-\frac{2}{2v+1}\frac{\alpha_g}{\alpha_c}\right]. \tag{12a}$$

The difference in received power on introducing and then removing the gas is then

$$\Delta P=8P_0\frac{\alpha_g}{\alpha_c}\left[\frac{v^2}{(2v+1)^3}\right]. \tag{12b}$$

ΔP attains its maximum value when

$$v=1 \quad \text{or} \quad R_c=r^2R_0.$$

Under this condition

$$\Delta P_{\max}=(8P_0/27)(\alpha_g/\alpha_c). \tag{12c}$$

Since ΔP must equal or exceed $4kTN\Delta f$, the minimum gas absorption which may be detected is

$$\alpha_{\min}=(27kTN\Delta f/2P_0)\alpha_c, \tag{13a}$$

or if we wish to express this quantity in terms of resonator Q

$$\alpha_{\min}=27kTN\pi\Delta f/\lambda P_0Q. \tag{13b}$$

A comparison between Eqs. (4) and (13a) shows that there is no significant difference between the resonant cavity method for detecting a small absorption coefficient and the matched guide method, provided that we design the systems in both cases to realize the maximum change in received power. If we maximize ΔP of Eq. (12b) and ΔP of Eq. (2), these power changes are equal if

$$Q_c/2=(27\pi/32e)s\simeq s, \tag{14}$$

where $Q_c=2\pi/\alpha_c\lambda$ and the matched guide is s half wave-lengths long. In short, a cavity whose Q has been halved by optimum coupling is as effective a gas cell for absorption measurements as a matched guide which is $Q_c/2$ half-waves long.

STARK MODULATION IN MATCHED GUIDE

We now proceed to analyze the Stark modulation method used in a matched wave guide system. At a pressure low enough to isolate absorption lines from each other, the absorption

coefficient in the neighborhood of frequency $\omega_0/2\pi$ is

$$G(\omega)=\frac{A^2\alpha_g}{(\omega-\omega_0)^2+A^2}, \tag{15a}$$

α_g is the power absorption coefficient when $\omega=\omega_0$, and $A/2\pi$ is the half-width of the line. In the second-order Stark effect ω_0 is given by

$$\omega_0=\omega_1+BE^2.$$

This is the type of effect encountered in the ammonia inversion spectrum. To be specific, for the ammonia (3,3) line the shift in an equipment like that described in reference (4) is approximately 15 Mc for a field of 1400 volts per cm, and the shift follows a square law. If we let

$$E=E_0\cos pt,$$

where $p/2\pi$ is the frequency of the applied Stark field,

$$G(\omega,\omega_s)=\frac{A^2\alpha_g}{[(\omega-\omega_1)-\omega_s\cos^2pt]^2+A^2} \tag{15b}$$

where

$$\omega_s=BE_0^2.$$

If we take as a measure of frequency the half-width of the line and let

$$n=(\omega-\omega_1)/A \quad \text{and} \quad m=\omega_s/A,$$

Eq. (15) becomes

$$g(m,n)=\alpha_g/[(n-m\cos^2pt)^2+1]. \tag{16}$$

After traversing a guide of effective length l, a field component at the receiver is

$$H=H_0e^{-\alpha_g l/2}\cos\omega t.$$

If the receiver is a power measuring device, such as a bolometer, or a crystal operating in a range for which it obeys a square law, power at the receiver is proportional to H^2. So in the following discussion we use the term detector current as synonymous with power.

$$2(H/H_0)^2=1-gl+\cos2\omega t-gl\cos2\omega t, \tag{17}$$

when $gl\ll1$.

The term gl is used to determine the frequency component in the detector output in the neighborhood of p, $2p$, and so on, while unity appearing in (17) gives the zero frequency current. The remaining terms give us components at twice the applied microwave frequency. Where we deal with a second-order Stark effect and do not apply a steady biasing field to the gas there are no components of $(H/H_0)^2$ at frequencies which are odd harmonics of $p/2\pi$ but only at even harmonics of this frequency. The amplitude of the component at $2p$ is given by the integral

$$C_2(m,n)=\frac{\alpha_g l}{\pi}\times\int_{-\pi}^{\pi}\frac{\cos x}{[n-(m/2)(1+\cos x)]^2+1}dx \tag{18a}$$

where

$$x=2pt,$$

but

$$[n-(m/2)(1+\cos x)]^2+1=(m^2/4)(\cos x+R)(\cos x+S),$$

where

$$R=1-(2n/m)-j(2/m),$$
$$S=1-(2n/m)+j(2/m).$$

The integrand in (17a) may be expanded in partial fractions to give

$$C_2(m,n)=\frac{4\alpha_g l}{m^2\pi(R-S)}\left[\int_{-\pi}^{\pi}\frac{\cos x\,dx}{\cos x+S}-\int_{-\pi}^{\pi}\frac{\cos x\,dx}{\cos x+R}\right], \tag{18b}$$

which, when evaluated, yields

$$C_2(m,n)=\frac{8\alpha_g l}{m^2(R-S)}\left[\frac{R}{(R^2-1)^{\frac{1}{2}}}-\frac{S}{(S^2-1)^{\frac{1}{2}}}\right]. \tag{18c}$$

Substituting the appropriate values for R and S we have

$$C_2(m,n)=\frac{2\alpha_g l}{m}\times\frac{[(m-2n)^2+4]^{\frac{1}{2}}}{[(n^2-mn-1)^2+(m-2n)^2]^{\frac{1}{2}}}\times\sin\left(\theta-\frac{\phi}{2}\right), \tag{18d}$$

where

$$\theta=\arctan[2/(m-2n)]$$

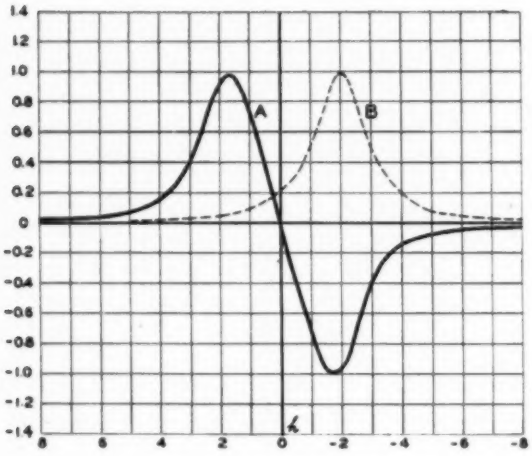


FIG. 3. Second harmonic in Stark modulation method.

and

$$\phi = \arctan[(m-2n)/(n^2-mn-1)].$$

This function is most conveniently studied by making the substitution $n=(m/2)+h$, since it is symmetrical about frequency h. Then

$$F_2(m,h) = \frac{4\alpha_0 l}{m} \frac{(h^2+1)^{\frac{1}{2}}}{\left\{\left[h^2-\left(\frac{m^2}{4}-1\right)\right]^2+m^2\right\}^{\frac{1}{4}}} \times \sin\left(\theta-\frac{\phi}{2}\right), \quad (19)$$

where

$$\theta = \arctan(-1/h)$$

and

$$\phi = \arctan\left[\frac{-2h}{h^2-(m^2/4)-1}\right].$$

$F_2(m,h)$ represents the amplitude of the detector current at the frequency p/π on a scale on which the amplitude of the direct current from the detector is unity. It may be shown that $F_2(m,h)$ attains a maximum value as a function of h when

$$h^2 = (m^2/4)-1. \quad (20)$$

Under these conditions

$$F_2(m,h) = \alpha_0 l(2(m-2))^{\frac{1}{2}}/m. \quad (21)$$

This function attains its maximum value, except for the physically meaningless value $m=0$, when $m=4$, and under these circumstances

$$F_2(m,h)_{\max} = \alpha_0 l/2. \quad (22)$$

The value of F_2 in (22) is quite uncritical as regards m, since the drop from the maximum value is only 3.5 percent when $m=3.2$ or 5.4. In short, to maximize the detector output the Stark shift in frequency should be four half-line widths.

Neglecting for the moment the modulation components of frequency $2p/\pi$, $3p/\pi$, and so on, we may write for the wave impinging on the detector the usual expression for a modulated wave

$$H = H_0(1+k\cos 2pt)\cos\omega t,$$

where k is the modulation factor used in the conventional radio engineering sense and here is $\alpha_0 l/4$. It is particularly to be noted from Eq. (22) that the change in power output as measured at the receiver is $\alpha_0 l$ when the optimum Stark field is used on a scale for which the steady received power is unity; moreover, this is precisely the power change for the general case given in Eq. (2). In short, the Stark modulation method is an effective one for making full use of the gas absorption.

The problem of measuring the absorption coefficient of the gas now involves nothing more than the simultaneous measurement of average received power, and of cyclic changes in received power, or what, in radio engineering, is known as the modulation factor.

The apparent shift in line frequency which results when one uses the Stark modulation method requires examination. A maximum receiver signal at frequency p/π occurs at two microwave frequencies given by

$$h^2 = (m^2/4)-1.$$

When $m=4$, these extremes occur when $n=2\pm\sqrt{3}$, but the position of the unperturbed line is given by $n=0$. In short, there is an apparent shift in line frequency of 0.27 half-line width if we adjust the Stark field, or the pressure, for maximum receiver output.

The detector output at four times the frequency of the Stark field is of interest and may be calculated using a procedure similar to that used in calculating the amplitude of the second harmonic. We now require the integral

$$C_4(m,n) = \frac{\alpha_0 l}{\pi}\int_{-\pi}^{\pi} \frac{\cos 2x\,dx}{(n-m\cos^2 x)^2+1}, \quad (23)$$

which, when evaluated, is

$$C_4(m,n)=\frac{8\alpha_0 l}{m^2}\left\{2-\frac{1}{R-S}\left[\frac{1-2S^2}{(S^2-1)^{\frac{1}{2}}}-\frac{1-2R^2}{(R^2-1)^{\frac{1}{2}}}\right]\right\}. \quad (24)$$

After making the substitution $h=n-(m/2)$ and some manipulation, (23) may be reduced to

$$F_4(m,h)=\frac{16\alpha_0 l}{m^2}\times\left[1+\frac{\left(\left[h^2-\left(\frac{m^2}{8}-1\right)\right]^2+\frac{m^2}{2}\right)^{\frac{1}{2}}}{\left(\left[h^2-\left(\frac{m^2}{4}-1\right)\right]^2+m^2\right)^{\frac{1}{2}}}\times\sin\left(\gamma-\frac{\phi}{2}\right)\right], \quad (25)$$

where

$$\gamma=\arctan\left[\frac{16h}{m^2-8(h^2-1)}\right]$$

and ϕ was defined under Eq. (19). F_4 represents the amplitude of the detector current at frequency $2p/\pi$ on the same scale as that used with $F_2(m,h)$. The maximum when $m=4$ in F_4 occurs at the point for which F_2 crosses the axis and has the value zero. At this point F_4 has the value $0.345\alpha l$. In short, under the above conditions the amplitude of the fourth harmonic current is 69 percent as large as that of the second harmonic but this peak is displaced from the position of the unperturbed line by $m/2$ half-widths. The information obtained from the fourth harmonic observations serves as a convenient check on that obtained from the second harmonic. F_4 asymptotically approaches zero as h becomes large as does F_2.

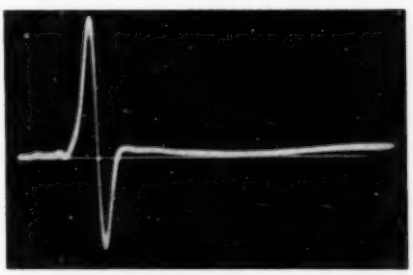


FIG. 4. Oscillogram of second harmonic.

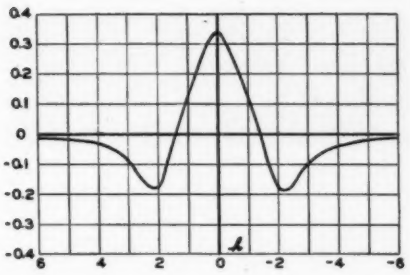


FIG. 5. Fourth harmonic in Stark modulation method.

A plot of Eq. (19) when $m=4$ is given in Fig. 3 while Fig. 4 is a photograph of the oscilloscope trace when the microwave frequency is swept through the ammonia (3,3) line at a 60-cycle rate. The two extremes are unequal in height in the photograph due to non-uniformity of the field in the gas chamber. Curve B gives the position of the unperturbed line while curve A is the plot of Eq. (19). A plot of Eq. (25), representing output at four times the frequency $p/2\pi$, for $m=4$ is shown in Fig. 5, while Fig. 6 is a photograph of the corresponding oscilloscope trace. The type of trace obtained depends in marked fashion on the value of m, that is, on the Stark frequency swing, and the photographs were obtained when m was slightly larger than 4.

The amplitude of harmonics of higher order may be calculated if they are of interest. Also, the alternating field may be superposed on a steady field, with the result that both even and odd harmonics of $p/2\pi$ will be found. The special case when the amplitudes of the two fields are equal has been used[4] and in this event, the results in Eqs. (19) and (25) are applicable to the frequencies $p/2\pi$ and p/π, respectively.

STARK MODULATION IN RESONANT CAVITY

In the present experiments the gas sample is enclosed in the annular space between two concentric brass cylinders approximately 6 feet long, the outer with an inside diameter of $5\frac{3}{4}$ inches and the inner with an outside diameter of 5 inches. The inner cylinder is made gas tight by means of two hemispheres soldered to its ends and is sup-

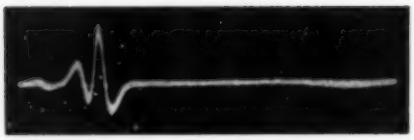


FIG. 6. Oscillogram of fourth harmonic.

ported and insulated by the use of six $\frac{3}{8}$-inch porcelain spacers. Thus a minimum amount of lossy material other than the gas is placed in the microwave field. A 100-kc Stark voltage is impressed on the inner electrode. With the cell geometry described, both the field and the field non-uniformity may be calculated readily when quantitative studies on the Stark effect itself are of interest rather than the detection of small lines. Input and output wave guides connected, respectively, to the microwave generator and the receiver are placed in opposite ends of the cylinder. Mica windows similar to those described by Malter[8] are used to make the system vacuum tight. Thus the cavity in use experimentally is a section of low loss wave guide but operating in a high order mode, and the results in Eq. (11) are applicable.

If, as in the matched guide, we impress an alternating Stark field of optimum value on the gas in the cavity, the absorption coefficient varies cyclically at frequencies near p/π, $2p/\pi$, etc. with amplitude of variation given by the Fourier coefficients in Eqs. (19) and (25). We are now in a position to calculate the modulation factor from Eq. (12) and from a knowledge of the modulation factor and Q of the cavity may infer the absorption coefficient.

A high modulation factor in itself is not enough to insure sensitivity in detecting weak absorption lines, if in the process of obtaining it we reduce the average power measured at the receiver. This unfavorable condition, namely, high modulation factor but low receiver power, is reached in a guide by the use of a length in excess of the optimum, and in the cavity by the use of coupling holes so small that the loaded Q is materially greater then $Q_c/2$ but the power through the cavity is low. There is no fundamental difference in the ability to detect small absorption coefficients, if, on the one hand, we use a matched guide of optimum length determined from α_c or, on the other hand, we use a short cavity with optimum coupling holes and whose Q is determined by the same α_c. Our choice of method will be based on convenience or technical reasons. In either event, the method of determining an absorption coefficient involves the measurement of a modulation factor.

The gas sample in a cavity is exposed to a microwave field more intense than that impressed on the sample in the matched guide. If saturation effects such as have been reported for ammonia are found,[7] the two methods will not yield concordant results. Moreover, the conclusion that the cavity and guide methods will, under optimum conditions, yield equivalent results is based on the assumption that they are constructed from guide material characterized by the same copper losses. When there is no need for an electrode for application of a Stark field, a more favorable cavity shape may be selected as in references 2 and 3 with a resulting increase in Q and marked improvement in the ability to detect feeble absorption lines.

The gas chamber in the present experiments is made up of a guide section having somewhat lower losses than that used by Hughes and Wilson and for this reason is capable of somewhat greater sensitivity.

CONCLUSIONS

We now come to a consideration of the minimum detectable absorption coefficient as given by Eq. (4) or (13),

$$\alpha_{\min}=\frac{4kTN\Delta f\epsilon}{P_0}\alpha_c. \tag{4}$$

Let us consider a numerical example such that $N=27$ db, $\Delta f=3000$ cycles, $P_0/e=10^{-10}$ watt, $\alpha_c=10^{-3}$ neper/cm.

Under these conditions the minimum detectable absorption is 2.5×10^{-7} neper/cm. The copper losses quoted are those measured for the copper guide commonly used at 24,000 Mc with inside dimensions 0.420 inch by 0.170 inch, while a 3000-cycle band width may be used if a 60-cycle sweep is employed with cathode ray tube indication. The calculated absorption of the (1,1) line of the inversion spectrum of N^{15} ammonia present in natural ammonia is 3.3×10^{-7} neper

[8] L. Malter, RCA Review **7**, 622 (1946).

per cm and detection of this line has been reported.[4]

The form of Eqs. (4) and (13) show the directions along which one may expect advances which permit the detection of smaller and smaller absorption lines. The effectiveness of crystal detectors arises from their low noise factor, but their utility is limited rather severely by their inability to make use of high microwave power. Moreover, they have the disadvantage that they operate as power measuring devices over but a limited range, so considerable care is required if one wishes to infer the modulation factor from measurements on the steady and alternating components of crystal current. To eliminate this difficulty, comparison experiments using a microwave signal generator with controlled and known modulation are required.

Increases in the power output of microwave generators will lead to advance only if such generators are themselves noise-free and if detectors with a favorable noise factor may be used with them. Some improvements may be possible by further reductions in guide losses, and the manner in which Q enters in Eq. (13) indicates the need for choice of low loss modes of operation.

The most striking improvements immediately attainable will be realized by a reduction of receiver band width, but such improvement is necessarily attended by an increase in the length of time it takes to explore a spectrum. The numerical example given was based on a 3000-cycle band width which is just adequate, if an absorption line is 1 Mc wide, to give us some knowledge of the absorption lines in a 50-Mc range sixty times per second. If we are content to sweep across a line, not in 1/3000 second, but in one-thirtieth second and use a recorder rather than an oscilloscope, the theoretical improvement in ability to detect small absorption lines is 100. The frequency range of klystrons in the 1.25-cm range is perhaps 5000 Mc, so the minimum time it would take to sweep over the range of such a tube is about three minutes.

Measurements of Dielectric Constant and Dipole Moment of Gases by the Beat-Frequency Method*

J. G. JELATIS
Laboratory for Insulation Research, Massachusetts Institute of Technology, Cambridge, Massachusetts

(Received January 9, 1948)

The beat-frequency method for measuring the dielectric constant and dipole moment of gases is critically investigated. It appears that the wide discrepancies in the data of previous observers are due in part, at least, to the neglect of certain stray components in the circuits involved. An apparatus taking these factors into consideration is described, and a method for the "absolute" standardization of the equipment is given. Results of dielectric constant and dipole moment measurements for several gases are presented and compared with previous data.

INTRODUCTION

THE equipment described in this paper was developed primarily to make possible measurements of the dielectric constant (κ') and dipole moment (μ) of gases. The design was, however, made flexible enough to permit its use wherever an apparatus is required for precisely measuring extremely small capacity variations.

The principle of operation of this instrument is not new, utilizing the beat oscillator method used by Zahn,[1] Smyth,[2] Watson,[3] and many others. Fundamentally this method measures the change in beat frequency between two oscillators when a test gas is introduced into a "gas con-

* This work was supported jointly by the Office of Naval Research and the Army Signal Corps, under ONR Contract N5ori-78 T.O.1.

[1] C. T. Zahn, Phys. Rev. **24**, 400 (1924).
[2] C. P. Smyth, *Dielectric Constant and Molecular Structure* (Chemical Catalog Company, New York, 1931).
[3] H. E. Watson, G. Gundu Rao, and K. L. Ramaswamy, Proc. Roy. Soc. **A132**, 569 (1931); H. E. Watson, Proc. Roy. Soc. **A143**, 558 (1934).

denser" forming a part of one oscillator. Alternatively, one may measure the capacitance change necessary to retune this oscillator to its original frequency. The dielectric constant is then calculated directly once the values of the fixed circuit parameters are known. By means of κ' measurements at different temperatures the dipole moment (μ) may be calculated according to Debye's treatment.[4]

Simple as the problem appears, it becomes apparent after a survey of the literature on the subject that considerable disagreement exists in the values obtained by various observers for the dielectric constant of many gases. These values differ generally by more than the claimed experimental errors allow. Hector and Woernley[5] in a recent paper set forth the facts admirably, including tabulations and analyses of the results of numerous observers.

It seemed, in view of this confused situation, that some supposedly minor design factors, responsible for the erratic data, were being neglected. A critical investigation of each component in the experimental set-up was in order. In addition, a method for the absolute standardization of such an apparatus must be used if really valid data are to be obtained. Complete success in these objectives is not claimed, but it is hoped some progress has been made.

GENERAL DESIGN

In Fig. 1 is shown a schematic block diagram of the particular system used, which, it must be made clear, offers no unusual advantages over the many other possible circuits, in all of which extreme stability and accurate calibration of circuit parameters are equally vital. The use of a commercial crystal frequency standard, the General Radio No. 475-C frequency monitor, for the fixed oscillator simply makes the task easier, as does the No. 661 frequency deviation meter for the beat frequency indicator. Since the latter is actually a frequency discriminator working about a mean of 1000 c.p.s., the variable oscillator need not be tuned closer than 1000 c.p.s. to the crystal standard to obtain a known beat frequency, thereby minimizing "pulling" effects between the oscillators. The variable oscillator output is taken from the untuned plate circuit allowing true "electron-coupled" operation, and thus contributing still further to circuit isolation and stability. The oscillator circuit is shown in Fig. 2a.

The temperature-controlled crystal standard is operated at 999 kc and the variable oscillator at 1000 kc, under which conditions the frequency discriminator shows a "zero-beat" or null indication. When the previously evacuated gas condenser (C_G) is filled with the test gas, the precision condenser (C_p) is varied until the frequency deviation meter once again gives a null reading. From the two readings of C_p, before and after the introduction of the gas, and a knowledge of the other circuit parameters, the dielectric constant may be calculated.

INDUCTANCE CORRECTION

It is in the consideration of these circuit parameters that most of the difficulty arises. The "simple" circuit of Fig. 2a would lead to a general result for $\kappa'-1$ as follows:

$$\kappa' C_G = C_G + \Delta C_G$$

$$\kappa' - 1 = \frac{\Delta C_G}{C_G}$$

where

C_G is gas cond. vacuum capacitance,
ΔC_G is variation due to gas of const. κ',
C_p'' is value of C_p without gas in C_G,
C_p' is value of C_p with gas in C_G,

$$= \frac{\dfrac{C_1 C_p''}{C_1 + C_p''} - \dfrac{C_1 C_p'}{C_1 + C_p'}}{C_G},$$

or when readings are reduced to S.T.P.

$$\kappa' - 1 = \frac{\dfrac{C_1 C_p''}{C_1 + C_p''} - \dfrac{C_1 C_p'}{C_1 + C_p'}}{C_G} \times \frac{760}{P_{\text{mm}}} \times \frac{T^0}{273.2}. \quad (1)$$

The *actual* circuit is considerably different,

[4] P. Debye, *Polar Molecules* (Dover Publications, New York, 1945), p. 36.
[5] L. G. Hector and D. L. Woernley, Phys. Rev. **69**, 101 (1946).
[6] H. E. Watson, Proc. Roy. Soc. **A143**, 558 (1934).

however, as Fig. 2b shows. Although most observers have been very careful to consider stray capacitances, the effect of residual inductances on measurements has been unexplainably neglected, and *may perhaps be the largest single factor leading to such a disparity of experimental results.* In Fig. 2b, the stray capacitances have been lumped into one, and may be disregarded since they do not enter into the calibration. The stray capacity that does exist across C_p and C_1 is important, but was taken care of by calibrating these condensers in their actual operating positions. The stray inductances, however, are not as easily disposed of, nor are they negligible in spite of the relatively low frequency used: in the present case, corrections for them amount to as much as 3 percent of $\kappa'-1$. If left uncorrected, this would introduce an error considerably larger than the next largest known probable error.

The effective capacity (C_{eff}) of a condenser is calculated from its direct current capacitance by the formula

$$C_{eff}=\frac{C_{d.c.}}{1-L\omega^2 C_{d.c.}}. \qquad (2)$$

The effective capacitance of the General Radio No. 722 precision condenser, to give an example, was calculated (for 1-megacycle freq.) from the given inductance value, and plotted as a function of the static capacity: at 1000 $\mu\mu$f the correction was 2.6 $\mu\mu$f, decreasing to negligible values at the low capacity end.

Watson,[6] it may be noted, mentions a perplexing effect (p. 559), namely, that he obtained higher ΔC readings for the same test gas at the high capacity end of a condenser than at the low end, but dismisses it without explanation. It seems likely that Eq. (2) would well have accounted for it and perhaps for many another "anomalous" effect. Watson did work out a minor correction for the inductance of a few connecting leads[6] (p. 565), and fixed condensers[3] (p. 574), but did not seem to feel the need for complete inductance correction of either the variable condensers or the gas condensers used.

The inductance of the gas condenser was calculated from measurements of its capacity at d.c. and at 1 megacycle against the precision condenser (taking into account the latter's residual inductance). But it is not sufficient to measure the $C_{G\,eff}$ at 1 mc and use that value in Eq. (1): as may be shown, this gives only partial correction. Complete correction is easily made as follows: since ΔC_G is generally very much smaller that C_G, it is permissible to differentiate Eq. (2), whence

$$\frac{d(C_{G\,eff})}{dC_G}=\frac{1}{(1-L_G\omega^2 C_G)^2},$$

or, using differentials,

$$\Delta(C_{G\,eff})=\frac{\Delta C_G}{(1-L_G\omega^2 C_G)^2},$$

which immediately gives the required correction factor. Actually, the inductance corrections for C_G and for C_p tend to neutralize each other, but in this particular case the net correction still amounted to approximately 2 percent.

The residual inductance of C_1, the series condenser, proved negligible. All capacitances were measured against the General Radio precision condenser, which itself was made as internally consistent as possible throughout the complete range. Its minimum capacitance was also checked against itself by substitution methods, so that ultimately all condensers in the system are calibrated in mutually consistent units; it is irrevelant whether this unit is exactly a micro-microfarad or not, since $\kappa'-1$ depends on the ratio of two capacitances.

An analysis of the inductive effects of the concentric line connecting C_G to the oscillator circuits showed them to be negligible for the length (18″) of line used. All other wiring in the oscillatory circuit was made as short as possible, since rigorous calculation of its inductive effects would be practically impossible.

STANDARDIZATION

It was due to the presence of such small but ever present factors of uncertainty that an independent standard for comparison and absolute calibration was felt vitally necessary.

After consideration of various possibilities, the use of the rare gases for this purpose was looked into. For them κ' is very nearly equal to the square of the optical refractive index, n^2 (Max-

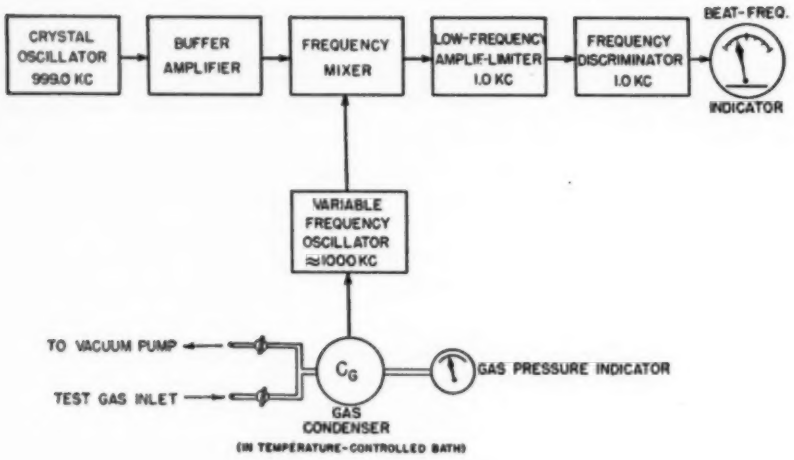


FIG. 1. Block diagram of system.

well's relation), the dispersion being very small between visible light and radiofrequencies, and calculable with negligible error.

Justification for using these gases is found in the extreme precision with which their refractive index can be measured by interferometric methods. Some of the most recent, and perhaps most valid of such measurements were made by C. and M. Cuthbertson[7] in 1932, with results as shown in Table I. The sixth and seventh columns give n^2 with the correction for dispersion at 1 mc; these are the values that should be obtained by direct κ' measurements. The optical values are, however, much more precise than the electrical measurements can be, and must be rounded off to at least one less figure than given when used for purposes of comparison. It is interesting to note that practically no differences exist in the values of the optical indices given by various observers when rounded off to the required accuracy; thus, the use of rare gases as a standard leaves little to be desired.

TABLE I. Refractive index, dispersion, and extrapolated static dielectric constant of the rare gases.[7]

Element	Early date (1910) $f_0^2\times10^{-27}$	Early date (1910) $C\times10^{-27}$	Latest data (1932) $f_0^2\times10^{-27}$	Latest data (1932) $C\times10^{-27}$	From early data $\kappa'-1$	From latest data $\kappa'-1$
Helium	34,991.7	2.4276	38,313.7	2.65228	0.00006930	0.00006923
Neon	38,916.2	5.18652	39,160.0	5.22606	0.00013327	0.00013345
Argon	17,008.9	9.43264	—	—	0.0005546	—
Krypton	12,767.9	10.6893	—	—	0.0008372	—
Xenon	8977.9	12.2418	—	—	0.0013635	—

(The above values to be substituted in the Sellmeier dispersion formula, $(\kappa'-1)=C/(f_0^2-f^2)$.)

[7] C. and M. Cuthbertson, Proc. Roy. Soc. **A135**, 40 (1932).

Watson *et al.*[3,6] seem to have been the first to make extensive comparisons between their electrical data on the rare gases and optical measurements. The excellent and consistent agreement of their values with the optical data, a fact quite singular among the many observers, lends added weight to the reliability of their measurements on other gases where such comparison is not possible.

SYSTEM SENSITIVITY AND ACCURACY

The over-all sensitivity, or smallest measurable external capacity change, is determined by the oscillator instability as the ultimate limit and by two additional factors: the accuracy with which the beat frequency can be reset and the minimum readable increment of capacity in the series combination of C_1 and C_p. The "frequency deviation meter" used permits beat-note determination to one cycle directly. The fixed series condenser C_1 was chosen so that a movement of the precision condenser C_p by one of its smallest subdivisions (0.2 $\mu\mu$f) near its high capacity end changes the beat note approximately one cycle; thus both factors reach their normal limit simultaneously. In terms of external capacity change, this direct reading limit represents 0.001 $\mu\mu$f; a more sensitive external meter and interpolation of the precision condenser dial allows an increase in sensitivity to about 0.0002 $\mu\mu$f before oscillator instability makes further increase impracticable. In terms of dielectric constant, the smallest detectable $(\kappa'-1)$ is thus 0.4×10^{-6}, $(\Delta C/C_G \approx 0.0002/500)$. This, it must be emphasized, is

not the "probable error" of a measurement, but rather the sensitivity limit.

The probable error is a function of this constant "reading" error plus the sum of numerous other *percentage* errors, including the gas condenser calibration accuracy of 0.1 percent, the temperature and pressure reading accuracy, each also about 0.1 percent, and the C_1, C_p series combination calibration, which careful analysis showed introduces a minimum uncertainty of 0.5 percent. The probable error of the final result is thus seen to be 100 percent for $(\kappa'-1)=0.000,000,4$, decreasing asymptotically to about 0.5 percent for the larger values of $(\kappa'-1)$. Figure 3 shows this error as a function of ΔC_p.

When measuring small $(\kappa'-1)$ values, by taking the mean of numerous observations the random "reading" error is greatly reduced, but the error resulting from the indeterminacy of the various parameters remains, except as noted later, under "Results."

When measuring dipole moments very little error (ΔT) in the measurement of T is allowable. Whereas in the case of κ' readings the fractional error was only $\Delta T/T°$ Kelvin, it may be shown that in this instance it becomes $2\Delta T/(T_2-T_1)$, where T_2 and T_1 are the extremes over which measurements are made.

METHOD FOR SIMPLIFIED CALCULATION OF DIELECTRIC CONSTANTS

With the inclusion of the inductance correction the expression for ΔC in Eq. (1) becomes

$$\Delta C=\frac{C_1\dfrac{C_p''}{(1-L_p\omega^2C_p'')}}{C_1+\dfrac{C_p''}{(1-L_p\omega^2C_p'')}}-\frac{C_1\dfrac{C_p'}{(1-L_p\omega^2C_p')}}{C_1+\dfrac{C_p'}{(1-L_p\omega^2C_p')}}. \quad (3)$$

The difficulties involved in making rapid calculations using this equation are obvious; but even with the use of seven-place log tables or automatic calculators, sufficient accuracy will not be obtainable. This is so because the difference between the two terms of Eq. (3) becomes evident only in the fourth or fifth significant figure for small $C_p''-C_p'$ readings.

The method evolved to overcome this difficulty consisted in differentiating the expression for the series combination,

$$C=\frac{C_1\dfrac{C_p}{(1-L_p\omega^2C_p)}}{C_1+\dfrac{C_p}{(1-L_p\omega^2C_p)}},$$

with respect to C_p, obtaining,

$$\frac{\Delta C}{\Delta C_p}=\frac{C_1^2}{[C_1(1-L_p\omega^2C_p)+C_p]^2}. \quad (4)$$

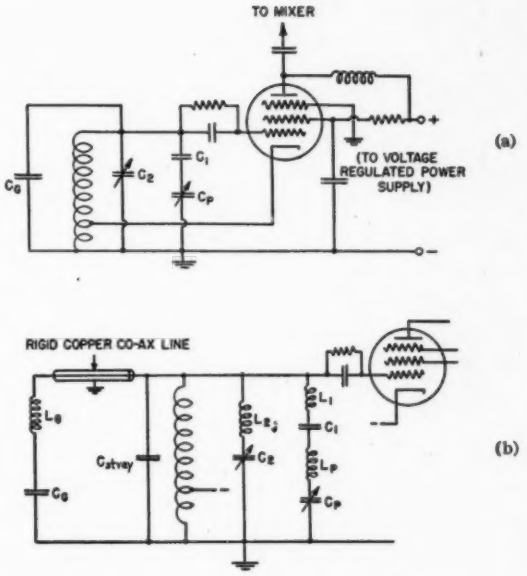


FIG. 2. a. Variable oscillator circuit — —. C_G = gas sample condenser, C_p = G. R. No. 722, 1000-$\mu\mu$f precision condenser, C_1 = 75-$\mu\mu$f zero temp. coeff. ceramic condenser, C_2 = shunt trimmer condenser. b. The same circuit, showing stray components.

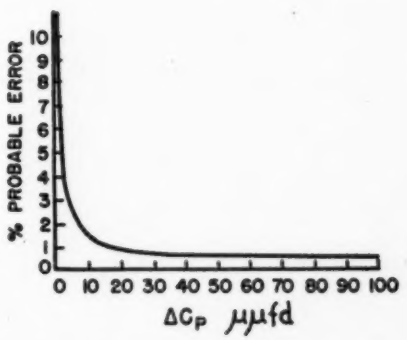


FIG. 3. Probable error *vs.* ΔC_p for a single observation, (based on 0.1-$\mu\mu$f random reading error, plus comprehensive system error of approximately 0.5 percent).

This equation was then used to calculate values of $\Delta C/\Delta C_p$ for every 10-$\mu\mu$f point in the upper 500-$\mu\mu$f range of the precision condenser C_p. These values ranged from approximately 0.004250 to 0.01330 $\mu\mu$f per $\mu\mu$f of C_p, and are represented with sufficient accuracy by only four significant figures. They were then plotted on a series of graphs with $\Delta C/\Delta C_p$ as ordinates and C_p as abscissas.

In addition, C is calculated very exactly from Eq. (3) for every integral 50-$\mu\mu$f value of C_p, and tabulated.

Reduction of the two readings, C_p'' and C_p', to dielectric constant is now very much simplified. If $C_p''-C_p'=\Delta C$ is smaller than 50 $\mu\mu$f, the mean value of $\Delta C/\Delta C_p$ is read from the graph (its slope changes very gradually), multiplied by ΔC_p, and the resultant ΔC substituted in Eq. (1) to give $(\kappa'-1)$ at S.T.P. conditions. When $C_p''-C_p'$ is larger than 50 $\mu\mu$f, the ΔC between 50-$\mu\mu$f intervals is read from the table mentioned above, while the residual is calculated as was done in the case of small $C_p''-C_p'$ differences, obtaining the total ΔC by addition.

This method has been in use constantly since its inception, and can reduce data in a few minutes that would, by "straightforward" calculations, have taken hours.

RESULTS

To date, dielectric constant measurements with this equipment have been made on helium, argon, neon, and oxygen. Dipole moment measurements include those of ethyl chloride (C_2H_5Cl) and carbon oxysulfide (COS), the first merely as a check against a much measured moment, and the latter as a verification of the moment obtained recently by an entirely independent method[8,9] that utilizes the Stark effect on the microwave spectrum of the gas.

TABLE II. Dielectric constant measurements.

	No. of observations	Mean $(\kappa'-1)\times10^7$ at S.T.P.	σ	$(\kappa'-1)\times10^7$ from Table I	Mean $(\kappa'-1)\times10^7$ Watson	Hector and Woernley[a]	Remarks
Helium	8	692	2	692.3	690[b]	684	Charcoal freshly activated
Helium	5	695	6	692.3	—	—	Charcoal left standing for several weeks
Helium	8	691	2	692.3	—	—	Charcoal freshly activated
Neon	6	1337	11	1334.5	1340[c]	1274	Large σ due to one unaccountably low reading
Neon	5	1341	3	1334.5	—	—	Same data, omitting above-mentioned reading
Argon	8	5542	9	5545.7	5504[c]	5451	—
Oxygen	6	5325	13	—	5300[b]	5233	—

[a] See reference 5.
[b] See reference 6.
[c] See reference 3.

TABLE III. Dipole moment measurements.

Gas	T_2-T_1	No. of observations at T_2	at T_1	$\mu\times10^{18}\pm$probable error
C_2H_5Cl	359°–292°K	5	11	(1.98±0.03) e.s.u.
COS	333°–265°K	3	5	(0.720±0.005) e.s.u.

Argon, neon, and oxygen were taken from the "spectroscopically pure" liter flasks now supplied by several concerns.[10] Although helium is also put out in this form, it could not be used because the flask pressure is under one atmosphere; whereas this pressure is sufficient in the case of the other gases, for helium two atmospheres was the lowest pressure that would cause sufficient capacity change in C_G (because of its unusually low κ') to be accurately measurable. This pressure caused a change in C_p of ≈12 $\mu\mu$f, which, from Fig. 3, is seen to represent approximately 1.5 percent error for a single observation; lower pressures would rapidly increase this value.

The helium was purified from ordinary "tank" helium by passing it through a charcoal filter immersed in a bath of liquid nitrogen. Considerable trouble was experienced in obtaining reproducible data until the charcoal trap had been baked out at about 500°C for a few hours, with a vacuum pump working continuously on it. Care was also exercised to avoid any exposure to air between the time of baking-out and making the helium measurements.

In Table II are given the condensed final results of these dielectric constant measurements and those of two other observers. Included are critical comparisons with optical data, except for oxygen. It will be noted that the difference between the mean value of $\kappa'-1$ and the extrapolated optical data is well within the accuracy of the present measurements as ex-

[8] T. W. Dakin, W. E. Good, and D. K. Coles, Phys. Rev. **70**, 560 (1946).
[9] M. W. P. Strandberg, *et al.*, to be published shortly.
[10] Linde Air Products Company, Tonawanda, New York, Air Reduction Sales Company, Stamford, Connecticut.

pressed by the standard deviation σ. The calibration of the condenser system may thus be deemed entirely satisfactory, and, in a sense "absolute," justifying the care taken in considering all the factors involved. The error curve (Fig. 3) now seems to have been definitely on the pessimistic side, and may, in the light of these measurements, be superseded by the standard deviation of any series of measurements. Any immediate, further increase in accuracy has to come from increased stability, and more refined gas-handling technique resulting in reduced σ values.

In the case of oxygen, comparison can be only with other observers' electrical data, since the optical measurements fall in a region of large dispersion. Once again particular reference must be made to Watson,[5,6] the only observer whose data closely and consistently parallel ours. His values for the rare gases are, in general, only slightly lower, and this correspondence holds also for oxygen. The results of Hector and Woernley (the most recent data, till now) show wide, unsystematic fluctuations.

Dipole measurements are recorded in Table III. A third temperature midway between T_1 and T_2 was used as a check point. The moment for ethyl chloride falls in among the many recorded values. The carbon oxysulfide value is considerably higher than the only other measurement (0.65×10^{-18} e.s.u.) made by this method;[11] it agrees quite exactly, however, with the moment obtained from the microwave Stark effect measurements (0.72×10^{-18} e.s.u.) by Dakin, Good, and Coles,[8] and somewhat less exactly with that of Strandberg, *et al.*[9] (0.739×10^{-18} e.s.u.) by the same microwave method.

Grateful acknowledgment is hereby made to Mr. T. Wentink of the M.I.T. Research Laboratory of Electronics for his kindness in preparing the carbon oxysulfide used in these measurements.

[11] C. T. Zahn and J. B. Miles, Jr., Phys. Rev. **32**, 497 (1938).

Here and There

New Appointments

Donald B. Harris has joined the Collins Radio Company of Cedar Rapids, Iowa, as Executive Assistant to the Director of Research.

M. N. Davis, Superintendent of the Physics Research Laboratory for the Kimberly-Clark Corporation since 1937, has been named Associate Technical Director. He will also continue as head of physics research. Dr. Davis was the first full-time physicist employed in the paper industry.

Effective the first of January, H. N. Swenson, E. M. Thorndike, and J. E. Woods became Associate Professors in the Department of Physics of Queens College, Flushing, New York.

Necrology

Alva W. Smith, Professor of Physics at The Ohio State University, died suddenly on January 21, 1948. He had been a member of its faculty for 33 years and was recognized as a teacher of unusual ability, grasping sympathetically both the intellectual and human interests, abilities, and needs of his students.

Frederic L. Bishop, Professor of Physics at the University of Pittsburgh and Dean of Engineering at that institution from 1909 to 1923, died October 11, 1947. He was nationally prominent as a physicist and was active in many governmental commissions during both world wars and the intervening period.

Congress of Applied Mechanics

The Seventh International Congress of Applied Mechanics will be held at the Imperial College of Science and Technology, South Kensington, London, England, September 5–11, 1948. The technical sessions will be organized under the following sections: elasticity and plasticity; aerodynamics, hydrodynamics, meteorology; thermodynamics, heat transfer, etc.; and vibrations, lubrication, and experimental methods. In addition, there will be several general lectures or surveys.

Arrangements have been made for visits to the National Physical Laboratory, the Royal Aircraft Establishment, the National Gas Turbine Establishment, and the General Electric Research Laboratories, during the week following the Congress.

Forms of application for membership and other material concerning the Congress can be obtained from the Organizing Secretary, Seventh International Congress of Applied Mechanics, Imperial College of Science and Technology, London, S. W. 7.

Joint Radio Meeting

The annual joint meeting of the American Section, International Scientific Radio Union, and the Institute of Radio Engineers will be held in Washington, May 3, 4, and 5, 1948. As usual, the program will be devoted to the more

fundamental and scientific aspects of radio and electronics. The program of titles and abstracts will be available in booklet form for distribution before the meeting. Anyone wishing to submit papers for presentation at this meeting should send title and a 100-word abstract as soon as possible to Dr. Newbern Smith, Secretary, American Section, URSI, National Bureau of Standards, Washington 25, D. C.

International Rubber Technology Conference

The Institution of the Rubber Industry, London, England, is planning another International Rubber Technology Conference, the first to be held since 1938. It will occur June 23–25, 1948, at the Central Hall, Westminster, London, S. W. 1. The Conference will cover the following subjects: natural and synthetic rubber latices, chemistry of rubber, physics of rubber, testing and analysis, synthetic rubbers, compounding ingredients, fibers, and textiles, developments in factory processes and products since 1938. Papers to be contributed should be addressed to the Institution of the Rubber Industry, 12, Whitehall, London, S. W. 1, England.

After the Conference, all the papers and full details of the Conference will be published in a volume entitled *The Proceedings of the Rubber Technology Conference, London, 1948*, a copy of which will be issued free to all those enrolling for the Conference. Extra copies will be available for sale generally at £2/2/0 each. The fee for enrollment is £2/2/0, and application forms and details can be obtained from the Conference Secretary at the Whitehall address above. The Secretary will arrange hotel accommodation to suit individual tastes and assist generally.

Foundry Congress

The 1948 Foundry Congress and Show of American Foundrymen's Association will be held in Philadelphia, May 3–7.

Lectures on Statistical Methods

The Massachusetts Institute of Technology has announced two series of lectures on statistical methods by L. H. C. Tippett, Chief Statistician, British Cotton Industry Research Association, and an address on the future of statistics in industrial research and quality control by Walter A. Shewhart of the Bell Telephone Laboratories. The lectures by Mr. Tippett will deal with statistical methods for industrial quality control and for technical investigation and experimentation, and will occur on May 5 through May 7, and May 12 through May 14 from 3–5 P.M. Mr. Shewhart's lecture will take place on May 14 at a dinner meeting. For further information write to Mr. H. A. Freeman, Massachusetts Institute of Technology, Boston 39, Massachusetts.

Symposium on Spectroscopic Equipment

The Society for Applied Spectroscopy, in cooperation with the Polytechnic Institute of Brooklyn, announces a Symposium on Spectroscopic Equipment, to be held at the Polytechnic Institute of Brooklyn on Saturday, May 22, 1948, under the chairmanship of Dr. W. L. Parker. There will be a morning and an afternoon session. The meeting will start at 9:30 A.M. The latest development on instruments in the field of absorption and emission spectroscopy will be exhibited by the leading manufacturers.

New Booklets

James G. Biddle Company, 1316 Arch Street, Philadelphia 7, Pennsylvania, offers its Bulletin 21–45, describing and illustrating the "Meg" type of insulation tester. 12 pages, free on request.

Allied Radio Corporation, 833 West Jackson Boulevard, Chicago 7, Illinois, announces the publication of its new 1948 172-page catalog, covering "everything in radio and electronics." Free on request.

The Rubber Chemicals Division of E. I. du Pont de Nemours and Company, Wilmington, Delaware, publishes a quarterly, *The Neoprene Notebook—Facts about Neoprene for the Engineer*. 8 pages.

Murex Limited, Rainham, Essex, England, has announced the publication of a leaflet entitled *Metal Powders for Welding Electrode Manufacture*. It is available in English, French, and Spanish. 4 pages, free on request.

The Atomic Energy Commission has published a series of lectures in the field of nuclear physics prepared by a group of outstanding experts who assisted in the development of the atomic bomb. These lectures were originally given at Los Alamos late in 1943 to aid in training personnel there in the fundamentals of nuclear science. The presentation consists of 41 chapters divided into six sections and is entitled *Lecture Series in Nuclear Physics*. It is available for 55 cents from the Superintendent of Documents, Washington 25, D. C.

Illinois Testing Laboratories, Inc., 420 North La Salle Street, Chicago 10, Illinois, has issued a leaflet entitled *Alnor Instruments*, its bulletin No. 2982. The company produces temperature and air velocity measuring and controlling instruments. 4 pages, free on request.

Sales Service Division, Eastman Kodak Company, Rochester 4, New York, has published another insertion for the *Kodak Photographic Notebook*. It is entitled "Kodachrome and Ektachrome exposure in photomicrography" and has 32 pages. Free on request.

General Radio Company, Boston 39, Massachusetts, has issued an 8-page pamphlet entitled *New Variac Continuously Adjustable Transformers*. Free on request.

American Instrument Company, Inc., of Silver Spring, Maryland, has published a 28-page bulletin on its line of electric hygrometer equipment for industrial and laboratory applications. The bulletin is numbered 2164–114 and is free on request.